University of Glasgow | Institute of Neuroscience & Psychology

Investigating the Neural Mechanisms Underlying Audio-visual Perception using Electroencephalography (EEG)

Stephanie Claire Boyle

A thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Neuroscience & Psychology

College of Medical, Veterinary & Life Sciences

University of Glasgow

February 2018

CENTRE FOR COGNITIVE NEUROIMAGING

BBSRC bioscience for the future

Abstract

Traditionally research into how we perceive our external world focused on the unisensory approach, examining how information is processed by one sense at a time. This produced a vast literature of results revealing how our brains process information from the different senses, from fields such as psychophysics, animal electrophysiology, and neuroimaging. However, we know from our own experiences that we use more than one sense at a time to understand our external world. Therefore to fully understand perception, we must understand not only how the brain processes information from individual sensory modalities, but also how and when this information interacts and combines with information from other modalities. In short, we need to understand the phenomenon of multisensory perception.

The work in this thesis describes three experiments aimed to provide new insights into this topic. Specifically, the three experiments presented here focused on examining when and where effects related to multisensory perception emerged in neural signals, and whether or not these effects could be related to behaviour in a time-resolved way and on a trial-by-trial basis. These experiments were carried out using a novel combination of psychophysics, high-density electroencephalography (EEG), and advanced computational methods (linear discriminant analysis and mutual information analysis).

Experiment 1 (Chapter 3) investigated how behavioural and neural signals are modulated by the reliability of sensory information. Previous work has shown that subjects will weight sensory cues in proportion to their relative reliabilities; high reliability cues are assigned a higher weight and have more influence on the final perceptual estimate, while low reliability cues are assigned a lower weight and have less influence (Angelaki, Gu, & DeAngelis, 2009; Fetsch, DeAngelis, & Angelaki, 2013a). Despite this widespread finding, it remains unclear when neural correlates of sensory reliability emerge during a trial, and whether or not modulations in neural signals due to reliability relate to modulations in behavioural reweighting. To investigate these questions we used a combination of psychophysics, EEG-based neuroimaging, single-trial decoding, and regression modelling. Subjects performed an audio-visual rate discrimination task where

the modality (auditory, visual, audio-visual), stimulus stream rate (8 to 14 Hz), visual reliability (high/low), and congruency in rate between audio-visual stimuli (± 2 Hz) were systematically manipulated. For the behavioural and EEG components (derived using linear discriminant analysis), a set of perceptual and neural weights were calculated for each time point. The behavioural results revealed that participants weighted sensory information based on reliability: as visual reliability decreased, auditory weighting increased. These modulations in perceptual weights emerged early after stimulus onset (48 ms). The EEG data revealed that neural correlates of sensory reliability and perceptual weighting were also evident in decoding signals, and that these occurred surprisingly early in the trial (84 ms). Finally, source localisation suggested that these correlates originated in early sensory (occipital/temporal) and parietal regions respectively. Overall, these results provide the first insights into the temporal dynamics underlying human cue weighting in the brain, and suggest that it is an early, dynamic, and distributed process in the brain.

Experiment 2 (Chapter 4) expanded on this work by investigating how oscillatory power was modulated by the reliability of sensory information. To this end, we used a time-frequency approach to analyse the data collected for the work in Chapter 3. Our results showed that significant effects in the theta and alpha bands over fronto-central regions occurred during the same early time windows as a shift in perceptual weighting (<100 ms and 250 ms respectively). Specifically, we found that theta power (4 - 6 Hz) was lower and alpha power (10 – 12 Hz) was higher in audio-visual conditions where visual reliability was low, relative to conditions where visual reliability was high. These results suggest that changes in oscillatory power may underlie reliability based cue weighting in the brain, and that these changes occur early during the sensory integration process.

Finally, Experiment 3 (Chapter 5) moved away from examining reliability based cue weighting and focused on investigating cases where spatially and temporally incongruent auditory and visual cues interact to affect behaviour. Known collectively as “cross-modal associations”, past work has shown that observers have preferred and non-preferred stimuli pairings. For example, subjects will frequently pair high pitched tones with small objects and low pitched tones with large objects (Parise & Spence, 2012; Spence, 2011). However it is still unclear

when and where these associations are reflected in neural signals, and whether they emerge at an early perceptual level or later decisional level. To investigate these questions we used a modified version of the implicit association test (IAT) to examine the modulation of behavioural and neural signals underlying an auditory pitch – visual size cross-modal association. Congruency was manipulated by assigning two stimuli (one auditory and one visual) to each of the left or right response keys and changing this assignment across blocks to create congruent (left key: high tone – small circle, right key: low tone – large circle) and incongruent (left key: low tone – small circle, right key: high tone – large circle) pairings of stimuli. On each trial, subjects were presented with only one of the four stimuli (auditory high tone, auditory low tone, visual small circle, visual large circle), and asked to respond which was presented as quickly and accurately as possible. The key assumption with such a design is that subjects should respond faster when associated (i.e. congruent) stimuli are assigned to the same response key than when two non-associated stimuli are. In line with this, our behavioural results demonstrated that subjects responded faster on blocks where congruent pairings of stimuli were assigned to the response keys (high pitch-small circle and low pitch-large circle), than blocks where incongruent pairings were. The EEG results demonstrated that information about auditory pitch and visual size could be extracted from neural signals using two approaches to single-trial analysis (linear discriminant analysis and mutual information analysis) early during the trial (<50ms), with the strongest information contained over posterior and temporal electrodes for auditory trials, and posterior electrodes for visual trials. EEG components related to auditory pitch were significantly modulated by cross-modal congruency over temporal and frontal regions early in the trial (~100ms), while EEG components related to visual size were modulated later (~220ms) over frontal and temporal electrodes. For the auditory trials, these EEG components were significantly predictive of single trial reaction times, yet for the visual trials the components were not. As a result, the data support an early and short-latency origin of cross-modal associations, and suggest that these may originate in a bottom-up manner during early sensory processing rather than from high-level inference processes. Importantly, the findings were consistent across both analysis methods, suggesting these effects are robust.

To summarise, the results across all three experiments showed that it is possible to extract meaningful, single-trial information from the EEG signal and relate it to behaviour on a time resolved basis. As a result, the work presented here steps beyond previous studies to provide new insights into the temporal dynamics of audio-visual perception in the brain. All experiments, although employing different paradigms and investigating different processes, showed early neural correlates related to audio-visual perception emerging in neural signals across early sensory, parietal, and frontal regions. Together, these results provide support for the prevailing modern view that the entire cortex is essentially multisensory and that multisensory effects can emerge at all stages during the perceptual process.

Table of Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Publications

Chapter 3:

Boyle, S. C., Kayser, S. J., & Kayser, C. (2017). Neural Correlates of Multisensory Reliability and Perceptual Weights Emerge at Early Latencies during Audio-visual Integration. *bioRxiv*, 116392. https://doi.org/10.1101/116392

Boyle, S. C., Kayser, S. J. and Kayser, C. (2017), Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. Eur J Neurosci, 46: 2565–2577. doi:10.1111/ejn.13724

Chapter 5:

Boyle, S.C., Ince, R.A.A., Kayser, C. (*in prep*). Neural Correlates of an Auditory Pitch - Visual Size Cross-modal Association Emerge Early during Perception.

Published work carried out alongside (but not included) in this thesis:

Lages, M., **Boyle, S. C.,** & Jenkins, R. (2017). Illusory Increases in Font Size Improve Letter Recognition. *Psychological science*, *28*(8), 1180-1188.

Lages, M., **Boyle, S.C.**, & Jenkins, R. (2015). Illusory Expansion Improves Visual Acuity. *Journal of vision*, *15*(12), 12-12.

Boyle, S. C., Jenkins, R., & Lages, M. (2014). Through the neural magnifying glass: visual acuity and motion-aftereffect. *i-Perception*, *5*(5), 479-479.

Lages, M., **Boyle, S. C.,** & Jaworska, K. (2013). Flipping a coin in your head without monitoring outcomes? Comments on predicting free choices and a demo program. *Frontiers in psychology*, *4*.

Abbreviations

AAL	Automated anatomical labelling
CRT	Cathode ray tube
Db	Decibel
EEG	Electroencephalography
EOG	Electro-oculogram
ERF	Event related field
ERP	Event related potential
fMRI	functional Magnetic Resonance Imaging
FLD	Fisher's linear discriminant
GCMI	Gaussian copula mutual information
HEOG	Horizontal electro-oculogram
IAT	Implicit association test
ICA	Independent component analysis
IPS	Intraparietal sulcus
LDA	Linear discriminant analysis
MEG	Magneto-encephalography
MI	Mutual information
MNI	Montreal neurological institute
PSE	Point of subjective equality
SC	Superior colliculus
SNR	Signal-to-noise ratio
SP	Spike potential
REOG	Radial electro-oculogram
ROC	Receiver operator characteristic
RT	Reaction time
SPL	Sound pressure level
STS	Superior temporal sulcus
TFR	Time-frequency representation
TPJ	Temporoparietal junction
TMS	Transcranial magnetic stimulation
VEOG	Vertical electro-oculogram

Acknowledgements

Many people have supported me during the past four years, and I wish to take a moment to thank them.

In terms of academic support, thanks are owed to many people. Thank you to the BBSRC for providing me with the funding to complete this work. Christoph, thank you for introducing me to an interesting Ph.D. topic, and for sharing your knowledge with me over the years. Joachim, thank you for the reassurance, support, and advice you gave over the final PhD months; you made the write up process seem much less daunting and I'm very grateful. Guillaume, thank you for the pre-viva pep talk, and for teaching me about open, robust science – there isn't a single bar graph in this thesis because of you. Finally, Martin, thank you for involving me in all of your cool projects throughout my Ph.D., and for helping me get my name on my first ever publication.

I would also like to say thanks to Phil, Helena, Heather, and Niamh, for encouraging me to teach and for giving me so many opportunities to do so. I learned a huge amount about teaching from all of you, and discovered a passion for teaching that I never knew existed.

Thanks are also owed to those who gave me academic advice: Anne, thank you for listening to my problems, and for kaffepausen with a side of Ph.D. advice. Bruno, thank you for all the motivational pep talks and academic mentoring; your "Bruno-isms" will not be forgotten!

Alex, Anne, Yulia, Robin, and Joachim, I owe you all a huge thank you for reading chapters of my thesis. This final version is much better thanks to all of you.

In terms of personal support, there are many people who deserve thanks. Merle and Alex, thank you for all the office beers, mountain adventures, memes, and endless advice. You both made the past four years very fun, and I'm glad we went through them together. Marc, Ewa, and everyone else I walked up a mountain with (yes, Lucy too), thank you for providing moral support and awesome company for much needed weekend escapes; our mountain faffing kept me sane. Yulia, Gemma, Matt, Kevin, and the rest of the PGR lot, thank you for assisting with experiments, good nights out, and lots of laughs.

Last but not least, I want to say a huge thank you to my family (Oscar and Maia included, of course). You've all supported me in so many different ways, and I appreciate everything you've done; thank you for being an awesome bunch. Dad, thank you for all the support and for keeping me sane with cycling days. Al, thanks for being a very supportive brother, and for all the pizza company. And Mum… thank you for being there every step of the way, supporting me no matter what, and believing in me when I didn't. Your encouragement and praise often fell on deaf ears, but it never went unappreciated. You helped me see that the light at the end of the "thesis tunnel" wasn't *really* that far away, and always inspired me to keep going. Thank you.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Stephanie Boyle

Chapter 1 : General Introduction

What is multisensory perception?

Multisensory perception is a term used to describe the integration and interaction of sensory information across different modalities. More specifically, multisensory integration refers to situations where two stimuli are perceived as one, and where the multisensory stimuli cause a significantly different neuronal response (e.g. increased neuronal firing) in multisensory compared to unisensory conditions (Stein & Meredith, 1993). In contrast, the term multisensory interaction describes situations where information from different sensory modalities is processed, but does not necessarily result in one percept and there is not a significantly different neural response (Calvert et al., 2012). Despite the distinct meanings, these two terms are often used interchangeably and without specific definition in the literature. However, the work in this thesis deals with cases of multisensory integration (where the stimuli are perceived as one – see Chapters 3 and 4) as well as multisensory interactions (where auditory information interacts with visual information – see Chapter 5), and so it is important to define the difference between the two.

When does multisensory perception occur?

Integration is most likely to occur when stimuli are spatially and temporally proximate (i.e. when they occur from the same place at the same time). This is best demonstrated by examining how behavioural responses are modulated during the presentation of two simultaneous stimuli. For example, perceptual sensitivity and response times for visual stimuli are enhanced when an auditory stimulus is presented at the same location (Bolognini, Frassinetti, Serino, & Làdavas, 2005) or same time (Vroomen & De Gelder, 2000). The detectability of a low intensity auditory tone is improved when it is paired with a simultaneous visual stimulus (Lovelace, Stein, & Wallace, 2003). Multisensory stimuli can even provide learning benefits: in one study, Seitz, Kim, & Shams, (2006) demonstrated that participants who received training with auditory and visual stimuli (motion in both modalities) showed greater learning of motion direction, both within and across training sessions. Finally, presenting three types of

sensory information (trimodal) has been shown to lead to greater improvement in reaction time than bimodal stimuli (Diederich & Colonius, 2004; Hecht, Reiner, & Karni, 2008a, 2008b). Thus, collectively the evidence demonstrates that spatially and temporally proximate stimuli lead to multisensory integration, and that combining information across modalities can provide perceptual benefits.

Integration between two slightly conflicting stimuli can also occur, as long as the stimuli are spatially and temporally congruent. One well-known example of this is called the McGurk effect; when presented with a simultaneous auditory "ba" and visual "ga", subjects will often report "hearing" the sound "da" (McGurk & Macdonald, 1976). This demonstrates a dynamic interaction whereby the brain combines conflicting information into a meaningful (but incorrect) percept, simply because the two pieces of information originated from the same place at the same time. Another famous demonstration of this effect comes from a set of seminal experiments using what is now known as the "two beep, one flash" paradigm (Shams, Kamitani, & Shimojo, 2000, 2002). In these studies, participants were presented with a visual disc that flashed only once, paired with one or two auditory beeps. In most cases, when the visual stimulus was accompanied by two beeps subjects reported "seeing" two visual flashes. This effect — where the rate of a visual stimulus is pulled towards the rate of the auditory — is known as temporal ventriloquism, and has been replicated by many other subsequent studies (Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Repp & Penel, 2002; Vroomen & De Gelder, 2004). In a similar way, a visual stimulus can affect judgements about an auditory stimulus. "Spatial ventriloquist" paradigms show that when visual and auditory stimuli are presented at the same time but at a spatial offset, the visual stimulus will "pull" the reported location of the auditory stimulus towards it (Bertelson, Vroomen, De Gelder, & Driver, 2000; Jean Vroomen, Bertelson, & de Gelder, 1998). Together, these effects show that the brain can flexibly adjust and integrate sensory information, even when there are slight temporal or spatial discrepancies.

Temporal and spatial ventriloquist paradigms also demonstrate another important part of multisensory integration: the process of sensory cue weighting. It is known that vision provides more accurate spatial information while audition

provides more accurate temporal information (Recanzone, 2003; Repp & Penel, 2002). Therefore in spatial ventriloquist tasks (where the visual modality is more "appropriate" for judging location), the visual information has greater influence on the final perceptual estimate and consequently pulls the auditory stimulus towards it. In temporal ventriloquist paradigms, the auditory modality is more appropriate to judge the temporal properties of stimuli, and so the rate estimate of the visual stimulus is pulled towards the auditory rate. Thus, not only is spatial and temporal congruency important for multisensory integration, so is the weighting of each cue. Investigating how spatially and temporally congruent and incongruent audio-visual stimuli are weighted and integrated is the focus of Chapters 3 and 4.

In contrast, multisensory interactions occur in a wider variety of contexts that do not necessarily depend on spatial and temporal congruence or perceptual relevance. For example, subjects show faster reaction times during stimulus discrimination when high pitched tones are paired with small visual objects, than when high pitched tones are paired with large objects (Bien, ten Oever, Goebel, & Sack, 2012; Evans & Treisman, 2010; Parise & Spence, 2012). Similarly, individuals will respond faster to visual stimuli in high positions of space when they are paired with high tones, compared to when they are paired with low tones (Evans & Treisman, 2010; Parise & Spence, 2012; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). These associations — known as "cross-modal correspondences" or "cross-modal associations" — have even been shown to occur when only a single stimulus is presented on a trial. In a recent paper, Parise & Spence (2012) demonstrated that subjects respond faster to preferred pairings of audio-visual stimuli (high tone – small circle and low tone – large circle) if they are assigned to the same response key than if they are not. Importantly, this occurred even though only a single stimulus was presented on a trial. Overall, these results show that sensory information from separate modalities can interact with one another even if there is no spatial and temporal congruency, and even if the stimulus in one modality has no predictive validity for the stimulus in the other. Chapter 5 investigates this issue further, by examining when an effect of a cross-modal interaction between auditory and visual stimuli emerges in neural signals.

How does multisensory perception occur in the brain?

Non-linear neuronal responses

Much of the early work into multisensory perception in the brain came from animal work. A landmark set of studies (for review, see Stein & Meredith, 1993), showed neurons in the superior colliculus (SC) of the cat displayed increased firing (enhancement), decreased firing (depression), or a response that was larger than the sum of the individual unimodal responses (superadditivity, considered by some to be a special case of enhancement) when presented with spatially and temporally congruent visual, auditory, and somatosensory stimuli (Meredith & Stein, 1983a; Meredith & Stein, 1986). Additionally, neurons demonstrated a principle known as "inverse effectiveness", where the strength of the multisensory response depended non-linearly on the strength of the stimuli: the more effective the individual unisensory stimuli, the weaker the multisensory response (Meredith & Stein, 1983a; Meredith & Stein, 1986). These observations led to three defining "rules" for multisensory integration at a neural level (although they also apply to behaviour, as previously introduced). "The spatial rule" states that integration is more likely to occur when the stimuli come from the same location in space. "The temporal rule" states that integration is more likely to occur when the stimuli arrive at the same time. Finally, "the law of inverse effectiveness" states that multisensory integration will be more likely when the unisensory stimuli are weak (Meredith & Stein, 1983b; Meredith & Stein, 1986).

Collectively, the results from neuroimaging suggest that multisensory interactions in the brain follow, at least to some extent, the effects observed in the neuronal responses in animal brains (Meredith & Stein, 1983b; Stein & Meredith, 1993). For example, perhaps the most common approach to examining multisensory perception with neuroimaging has been to examine whether the multisensory neuronal response is significantly different from the combined sum of unisensory responses (following the guidelines of superadditivity seen in neuronal responses). Using this approach, a nonlinear response is calculated as: AV = AV-[A+V], where AV represents the audio-visual response, A represents the auditory response, and V represents the visual. Such a model has been successfully used to show multisensory interactions emerging in neural signals in

both EEG (Fort, Delpuech, Pernier, & Giard, 2002; Foxe et al., 2000; Giard & Peronnet, 1999; Molholm et al., 2002; Möttönen, Schürmann, & Sams, 2004) and fMRI work (Calvert, Campell, & Brammer, 2000; Calvert, Hansen, Iversen, & Brammer, 2001; Stevenson, Kim, & James, 2009). The principles of inverse effectiveness also seem to hold in such large scale neural activity. For example, a study by Senkowski, Saint-Amour, Höfle, & Foxe, (2011) showed that the strongest amplitude modulations in the multisensory EEG signal occurred when the respective auditory and visual unisensory stimuli were of low intensity. Crosse, Di Liberto, & Lalor, (2016) found similar representations of inverse effectiveness in EEG signals in response to audio-visual speech. Finally, recent functional magnetic resonance studies (fMRI) have reported findings in line with the principle of inverse effectiveness occurring in the superior temporal sulcus (STS) (Stevenson & James, 2009; Werner & Noppeney, 2010) and nearby cortical regions (Holle, Obleser, Rueschemeyer, & Gunter, 2010) during presentation of audio-visual stimuli containing varying degrees of noise.

However, some have cautioned against relying too heavily on the principles observed at the level of the single neuron when making inferences from neuroimaging (Beauchamp, 2005; Calvert, 2001; Giard & Besle, 2010; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). In neuroimaging, the signal being measured is not from a single neuron, but rather is the composite activity from millions of neurons. Consequently, any information recorded from the scalp will be a mix of activity and noise from many sources in the brain. This creates several problems when looking for non-linear responses in neuroimaging signals using [AV-(A+V)], as it assumes that: (a) the auditory and visual signals contain no overlapping information; (b) the activity in either unisensory condition is not modulated by the presence of another stimulus; and (c) biases — such as attentional demands — are the same in both modalities you are measuring. Assumption (a) is easily violated if an electrode measuring visual activity contains any shared activity that an auditory electrode does. As each electrode measures a combination of neuronal responses, noise, and non-cognitive related activity from multiple sources in the brain this assumption is most likely violated. From research, we know that assumption (b) is also easily violated: for example, research has shown that auditory cortex responses can be modulated by visual stimuli (Kayser, Petkov, & Logothetis, 2008; Laurienti et al.,

2002; Vetter, Smith, & Muckli, 2014). Finally, assumption (c) could be easily violated if, for example, the behavioural task was a spatial localisation paradigm and it was necessary to pay more attention to the visual than auditory modality. Therefore, while non-linear effects have been found widely in neuroimaging, it is also important to examine and test other models of cue combination.

Linear cue combination

"Optimal Integration" or "Bayesian Integration" models are frameworks based on probability theory, which describe how cue integration can occur in a linear, statistically optimal fashion by following the rules of Maximum Likelihood Estimation (Angelaki et al., 2009; Rohde, van Dam, & Ernst, 2015). Under this framework, the "optimal" way to derive the most precise perceptual estimate is to generate a weighted average of the unimodal estimates, where each sensory cue is weighted relative to its individual precision and reliability. This weighting strategy is important, as the information coming in through our senses is inherently probabilistic; sensory cues can be unreliable (e.g. visual stimuli become less reliable in dim conditions) or redundant (e.g.an auditory stimulus might provide no additional information about the location of an event yet it occurs simultaneously with the visual). Alongside this, internal noise (e.g. in the response of neurons, muscle movements, or heartbeats) can create more uncertainty. This uncertainty lends itself well to using probability theory (i.e. Bayesian/optimal integration) models to investigate multisensory integration in the brain.

These frameworks have two predictions. First, perceptual estimates should be more precise for multisensory versus unisensory stimuli. Second, as the reliability of a sensory cue decreases, so should the weight assigned to it. For example, imagine you are performing an audio-visual spatial localisation task where the reliability of the auditory cue varies from low to high quality. According to the Optimal Integration framework, on trials where the visual and auditory stimuli are of equal reliability, your estimate of the location of the stimuli should be based on an equal weighting of both stimuli. However, on trials where the visual stimulus is of high quality and the auditory stimulus is of low quality, your estimate of the location of the audio-visual stimulus should be

based on unequal weighting; specifically, the visual modality should be weighted higher as it is of greater reliability.

These models have had great success in explaining cue integration in a variety of behavioural tasks. For example, Sheppard, Raposo, & Churchland, (2013) investigated how subjects weighted sensory cues during an audio-visual rate discrimination task. Subjects (humans/rats) were presented with two sequential streams of visual, auditory or audio-visual flicker, and had to judge which stream had a higher stimulation rate. On some audio-visual trials the reliability of the visual or auditory stimulus varied and — importantly — a discrepancy was introduced between the auditory and visual cues which caused the stimuli to flicker at slightly different rates. This incongruency in cue weighting paradigms is essential, as it allows assessment of the degree to which subjects are biased towards each cue: if subjects' responses are more in line with of one modality, it suggests they are placing higher importance on it. In this study, this bias was quantified by fitting psychometric curves to the behavioural response data to generate a set of "perceptual weights" for each modality. Indeed, the results revealed that for both rat and human subjects, performance was higher in the multisensory condition compared to the unisensory condition, and there was increased perceptual weighting for the more reliable stimuli compared to the less reliable. Similar findings were reported in a study by Alais & Burr, (2004), who investigated the integration of audio-visual cues during a spatial localisation task. Consistent with the predictions from optimal integration models, their results showed that localisation in the bimodal conditions was better than localisation in either the visual alone or auditory alone condition, and that when the reliability of the visual information decreased the auditory weighting increased. Overall, these findings of cue weighting have been demonstrated in many other behavioural studies, suggesting it a robust and important process during multisensory integration (Battaglia, Jacobs, & Aslin, 2003; Butler, Smith, Campos, & Bülthoff, 2010; Ernst & Banks, 2002; Fetsch, Turner, DeAngelis, & Angelaki, 2009; Helbig & Ernst, 2007; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Rosas, Wagemans, Ernst, & Wichmann, 2005; Sheppard, Raposo, & Churchland, 2013).

These principles of cue weighting have also been shown to emerge in neural signals. For example, single-cell recordings have shown that "neural weights"

(calculated via fitting psychometric curves to neural data) can vary with cue reliability in a manner consistent with predictions from statistical optimal models (Gu, Angelaki, & Deangelis, 2008; Morgan, DeAngelis, & Angelaki, 2008). Additionally, neural weights can also can predict the perceptual weights derived from behaviour (Fetsch, Pouget, DeAngelis, & Angelaki, 2012). Similarly, fMRI studies have demonstrated that BOLD responses are modulated by sensory reliability during visual-tactile (Beauchamp, Pasalar, & Ro, 2010; Helbig et al., 2012) and audio-visual tasks (Rohe & Noppeney, 2016) in a manner consistent with cue weighting frameworks: as the reliability of a stimulus decreases, the BOLD signal tends to increase in areas underlying the processing of the more reliable stimulus. Finally, recent fMRI work has indicated that this sensory weighting process emerges gradually along the cortical hierarchy, with stimulus reliability encoded in low-level sensory regions and cue weighting occurring in higher-order parietal regions (Rohe & Noppeney, 2015a, 2016). Overall, these results indicate that linear cue combination can be implemented by neurons in the brain, and that these computations can be measured by neuroimaging. Chapters 3 and 5 of this thesis use such a linear modelling to study the neural underpinnings of audio-visual cue perception across two different audio-visual tasks.

Neural oscillations

The final method (relevant to this thesis) by which the brain has been proposed to combine sensory information is via oscillatory activity. It has been known for almost a century that the brain shows rhythmic activity, and this activity shows both correlational and causal relationships with behaviour (Buzsáki & Draguhn, 2004; Wang, 2010). These rhythms – known as "neural oscillations" – are usually sub-divided by boundaries into separate bands based on the frequency with which they oscillate: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (>30 Hz). However it is important to remember that these frequency boundaries are somewhat arbitrary and can change depending on the specific experiment.

Importantly, oscillatory activity has been widely observed during multisensory tasks. For example, Sakowitz, Quiroga, Schurmann, & Basar, (2001) found increased theta activity in response to audio-visual stimuli compared to

unisensory stimuli. Similarly, Sakowitz, Quiroga, Schurmann, & Basar, (2005) found oscillatory power changes in theta and alpha evident early after presentation of audio-visual stimuli. Another study found increased beta band activity in response to audio-visual stimuli, and demonstrated that power also predicted shorter reaction times for multisensory stimuli versus unisensory stimuli (Senkowski, Molholm, Gomez-Ramirez, & Foxe, 2006). Activity in higher frequency bands is also a common finding in multisensory research: increased gamma activity has been found for congruent versus incongruent pairs of audio-visual stimuli (Schneider, Debener, Oostenveld, & Engel, 2008; Yuval-Greenberg & Deouell, 2007), and in response to mismatches in congruency between auditory and visual speech (Arnal, Wyart, & Giraud, 2011). Findings of oscillatory activity in multisensory paradigms are in fact so widespread that it has now been proposed that oscillations may be the key mechanism by which the brain combines information across different modalities (Engel, Senkowski, & Schneider, 2012; Fries, 2005). Under such a hypothesis, synchronised oscillations between different areas of the brain might provide the potential mechanism for crossmodal integration (Engel, Fries, & Singer, 2001), and modulations in power of phase of neural signals may reflect ongoing modulations or interactions due to multisensory perception (for review see Engel et al., 2007). Chapter 4 builds on this literature by investigating how oscillatory power is modulated by the reliability of multisensory stimuli.

Where and when do multisensory effects occur in the brain?

The final aspect of multisensory perception to review relates to *where* and *when* it occurs in the brain. The traditional view of multisensory processing was that it occurred in a hierarchical manner, with incoming sensory information first processed in early sensory areas or specialised subcortical regions, and later integrated at higher association areas (Meredith & Stein, 1986b; Stein & Meredith, 1993). And indeed, multisensory effects have been found widely in subcortical regions such as the superior colliculus (SC), basal ganglia and putamen (Stein & Meredith, 1993). With regards to cortical areas, in both animal and human studies, effects of multisensory perception have been found in the superior temporal sulcus (STS) (Beauchamp, Lee, Argall, & Martin, 2004;

Beauchamp, Yasar, Frye, & Ro, 2008; Benevento, Fallon, Davis, & Rezak, 1977; Bruce, Desimone, & Gross, 1981), parietal cortex (Avillac, Hamed, & Duhamel, 2007; Bremmer, Klam, Duhamel, Hamed, & Graf, 2002; Rohe & Noppeney, 2015a, 2016), and frontal and prefrontal areas (Fuster, Bodner, & Kroger, 2000; Senkowski et al., 2006). Thus, it appears there is a lot of evidence indicating multisensory integration occurs at specialised subcortical, association, and higher-order regions.

Yet in more recent years accumulating evidence has challenged this view (Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006). Many areas that were typically considered unisensory also show responses to multisensory stimuli. For example, animal work has demonstrated both somatosensory (Kayser, Petkov, Augath, & Logothetis, 2005; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007) and visual stimulation (Bizley & King, 2009; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Kayser, Petkov, Augath, & Logothetis, 2007) can modulate activity in early auditory cortex. At a more macroscopic level, it has been shown that event related potentials (ERP) and event related fields (ERF) in auditory cortex can be modulated by visual stimuli (Besle, Fort, Delpuech, & Giard, 2004; Colin et al., 2002; Möttönen, Krause, Tiippana, & Sams, 2002), and in somatosensory cortex in response to auditory-somatosensory stimulation (Foxe et al., 2000). Functional MRI studies have also revealed that visual stimuli can modulate auditory cortex activity as measured by the BOLD signal (Calvert, 2001; Calvert et al., 1997; Lehmann et al., 2006; Pekkola et al., 2005, 2006), and that sound content can be decoded from early visual areas (Vetter et al., 2014). Based on all this evidence, the modern view is now that early sensory areas are capable of processing stimuli from multiple modalities, thus rendering the cortex essentially multisensory.

Multisensory interactions have also been shown to occur surprisingly early in the perceptual process. For example, enhancement of auditory ERPs have been found in response to a simultaneously presented somatosensory stimulus as early as 50 ms after stimulus onset (Murray et al., 2005). Another study showed audio-visual ERP responses emerging at 46 ms after presentation in early parietal and occipital regions (Molholm et al., 2002). Fort et al., (2002) also demonstrated early (<200 ms) effects occurring in early sensory cortex during audio-visual object recognition. Overall, the early onset of these results suggests

multisensory effects cannot possibly be restricted to a late stage of the perceptual hierarchy.

However, much of the previously cited work relied on measuring the timing or fluctuation in peaks of ERPs, potentially due to the ease of which ERPs can be calculated, and/or the feasibility in collecting vast number of trials for analyses, difficulty implementing single-trial analysis classifiers, and longer computation time needed for single-trial analysis. While informative in their own right, such studies limit the amount of information one can draw from these early modulations. Thus, more work using a larger variety of tasks and methods is needed to examine when multisensory effects emerge in neural signals, and to further resolve what role early and low level sensory effects play in multisensory perception. This issue is a core topic of this thesis, and is investigated in Chapters 3, 4 and 5.

Chapter 1: Summary and Thesis Rationale

To summarise, multisensory perception is an important part of everyday life, it modulates behavioural and neural responses on a variety of levels and in many different ways, and research points towards the idea that the cortex is essentially multisensory. However, we still do not have a full understanding of how and when effects of multisensory perception emerge in the brain during a variety of tasks, or how modulations in neural signals due to multisensory perception relate to human behaviour. The work in this thesis aimed to provide insights into these open questions. Specifically, I carried out three experiments that investigated when effects related to multisensory perception emerged in neural signals, and examined whether these could be related to behaviour. To do so, we used a combination of psychophysics, high-density electroencephalography (EEG), and advanced computational methods (linear discriminant analysis and mutual information analysis). This novel combination made it possible to extract information from the EEG signal and relate it to behaviour on a trial-by-trial and time resolved basis. This steps beyond previous studies in the field, which have traditionally relied on examining multisensory perception via fluctuations in the amplitude or timing of neuronal responses, or by looking for increases and decreases in averaged voxel activations as measured

by fMRI. Thus, the work in this thesis provides novel insights into the temporal dynamics underlying audio-visual perception.

Thesis at a Glance (Abstracts)

Chapter 3: Neural Correlates of Multisensory Reliability and Perceptual Weights Emerge at Early Latencies during Audio-visual Integration

To make accurate perceptual estimates observers must take the reliability of sensory information into account. Despite many behavioural studies demonstrating that subjects weight individual sensory cues in proportion to their reliabilities, it is still unclear when during a trial neuronal responses are modulated by the reliability of sensory information, or when they reflect the perceptual weights attributed to each sensory input. We investigated these questions using a combination of psychophysics, EEG based neuroimaging, and single -trial decoding. Our results show that the weighted integration of sensory information in the brain is a dynamic process; effects of sensory reliability on task relevant EEG components were evident 84 ms after stimulus onset, while neural correlates of perceptual weights emerged 120 ms after stimulus onset. These neural processes had different underlying sources, arising from sensory and parietal regions respectively. Together these results reveal the temporal dynamics of perceptual and neural audio -visual integration and support the notion of temporally early and functionally specific multisensory processes in the brain.

Chapter 4: Theta and Alpha Power are modulated by Sensory Reliability Early during Audio-visual Integration.

Oscillations have been found in a variety of multisensory tasks, and are thought to be a fundamental process by which the brain combines information across the senses. However, it has yet to be investigated whether oscillatory power is modulated by the reliability of sensory information. Here we used a time-frequency approach to investigate whether the oscillatory power underlying auditory signals was modulated by changes in visual reliability during audio-visual integration. Additionally, we examined whether any changes in oscillatory

power occurred at the same time as changes in perceptual weighting. Our results revealed that auditory power during audio-visual conditions is modulated by visual reliability early during the perceptual process. Specifically, we found significantly lower theta power and higher alpha power during audio-visual conditions where the visual reliability was low, relative to conditions where visual reliability was high. These modulations in power occurred early in the trial from stimulus onset to 150 ms, and from 252 ms to 300 ms for theta and alpha respectively, over fronto-central sites. Together these results suggest a role of theta and alpha in sensory reweighting, occurring at an early stage during audio-visual integration.

Chapter 5: Neural Correlates of an Auditory Pitch – Visual Size Cross-Modal Association Emerge Early during Perception

Humans have been shown to exhibit implicit perceptual associations across the different senses. For example, subjects often associate high pitch tones with small objects and low pitch tones with large objects. However, it is remains unclear when and where these cross-modal associations are reflected in neural signals, and whether they emerge at an early perceptual level or later decisional level. In this study we investigated these questions using a modified version of an implicit association task (IAT) combined with 128-channel EEG and two approaches to single trial analysis. Our behavioural results revealed subjects identified stimuli faster during congruent than incongruent blocks, even though only a single stimulus was presented per trial. The EEG results demonstrated that EEG components related to acoustic pitch and visual size were early in the trial (~100 ms,~200 ms respectively), over temporal and frontal regions. For the auditory trials, these EEG components were significantly predictive of single trial reaction times, yet for the visual trials the components were not. As a result, our data support an early and short-latency origin of cross-modal associations, and suggest that these may originate in a bottom-up manner during early sensory processing rather than from high-level inference processes. Importantly, our EEG results were consistent across both EEG analysis methods, demonstrating that our findings are robust to the statistical methodology used.

Chapter 2 : Methods Overview

Electroencephalography

Electroencephalography (EEG) is a non-invasive brain imaging technique that records electrical activity from the brain. To record the EEG signal, electrodes are placed upon the scalp and these pick up the electrical current arising from the synchronised firing of millions of neurons (see Figure 2.1). These electrodes are placed at multiple sites on the scalp, which gives many signals to work with (typically >64 channels) and allows recording of signals from the whole brain simultaneously. Additionally from a single electrode, different measures such as amplitude, latency, phase and frequency, can be extracted. This provides at least four dimensions to work with. Finally, signals from electrodes are typically recorded thousands of times per second, resulting in a multi-dimensional signal with high temporal resolution. This is the main benefit of using EEG to study cognitive processes, as neural responses are modulated by internal and external events on the order of milliseconds rather than seconds.

The main limitation with EEG concerns spatial accuracy: EEG has a large spatial scale on the order of cm, as each electrode records activity not only from neurons directly below it, but also from the surrounding electrodes and other distributed sources in the brain. Additionally, the electrical propagation properties of the EEG signal mean that signals radiate out and away from the sources in different manners depending on the orientation of the neurons and the tissues they travel through. These factors do not mean that EEG cannot be used to study *where* a process occurs in the brain, but they do add a layer of difficulty to spatial interpretations.

With regards to the work contained in this thesis, the timing of multisensory perception in the brain was the focus of interest, and functional localisation was not of the upmost importance; as a result, EEG was the preferred recording method.

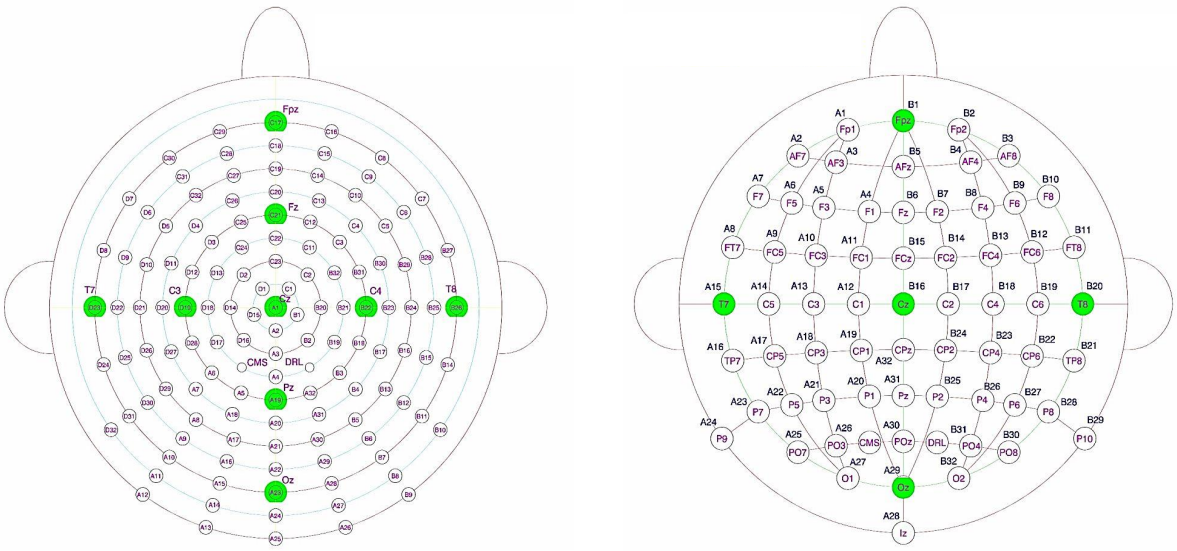


Figure 2.1 | EEG Biosemi Setup. Schematic representations of 128 channel (left) and 64 channel (right) systems used for the work in this thesis. Top represents the front of the head, with left and right on the corresponding sides.

Preprocessing Methods

As described above, the raw EEG recording consists of one time continuous signal per electrode composed of a mix of signal and noise. In this context, "noise" broadly refers to any signal — external or internal — that is not related to a cognitive process. External noise can arise from electrical devices (such as mobile phones, computers, and overhead lights), or recording equipment (e.g. broken electrodes). Internal sources of noise arise from sources such as eye movements, eye-blinks, heartbeat, body movements, and neural processes related to phenomenon other than the task. In order to disentangle these noise artifacts from true, task-relevant signal, preprocessing methods are used.

For the data contained in this thesis, preprocessing involved epoching data so that individual trials were separated and removing trials with high amplitudes (±120 μV) fluctuations. In addition to this, artefactual trials related to eye blink or eye movements were identified using independent components analysis (ICA) and removed using a combination of automatic and visual procedures. As some of these automatic removal processes go beyond "standard" techniques, and as the same preprocessing methods were used for the data in every chapter, an

overview of each removal technique will be described here once. However, experiment specific parameters (e.g. time window lengths, epoch lengths, etc.) will be described within each relevant chapter.

Artifact identification based on Independent Component Analysis (ICA)

Independent component analysis (ICA) is a statistical technique that uses linear decomposition to separate the EEG signal into a set of independent components. Using it as a method of artifact removal is based on the assumptions that the activity recorded on the scalp is a mix of independent sources related to artifacts and brain activity, and that these signals propagate linearly out from the sources in the brain. Thus, if the independent sources in the data can be identified using ICA, the ones related to artifacts can be removed.

ICA works by finding a set of weights (W) that linearly decompose the EEG signal (X) into a set of independent components, which provide information about the time course and spatial topography of the signal. The time course can be examined to find large amplitude fluctuations in the data, which typically represent noisy channels or eye blinks. The spatial topography of the components can be examined to see whether they bear typical artifact topographies (Delorme, Makeig, & Sejnowski, 2001; Li, Ma, Lu, & Li, 2006; Viola et al., 2009). For example, components showing strong activity centred over frontal or temporal electrodes are usually considered to be artifacts due to the close proximity of these electrodes to the eyes and the ears. Strong activity centred over a single electrode is also usually considered an artifact, as it is very unlikely that you would see EEG activity at only one electrode and not at neighbouring ones. Figure 2.2 provides some example topographies showing such artifacts.

Once these artifact components have been identified they can be removed and the remaining components can be projected back onto the scalp channels to produce artifact free EEG data.

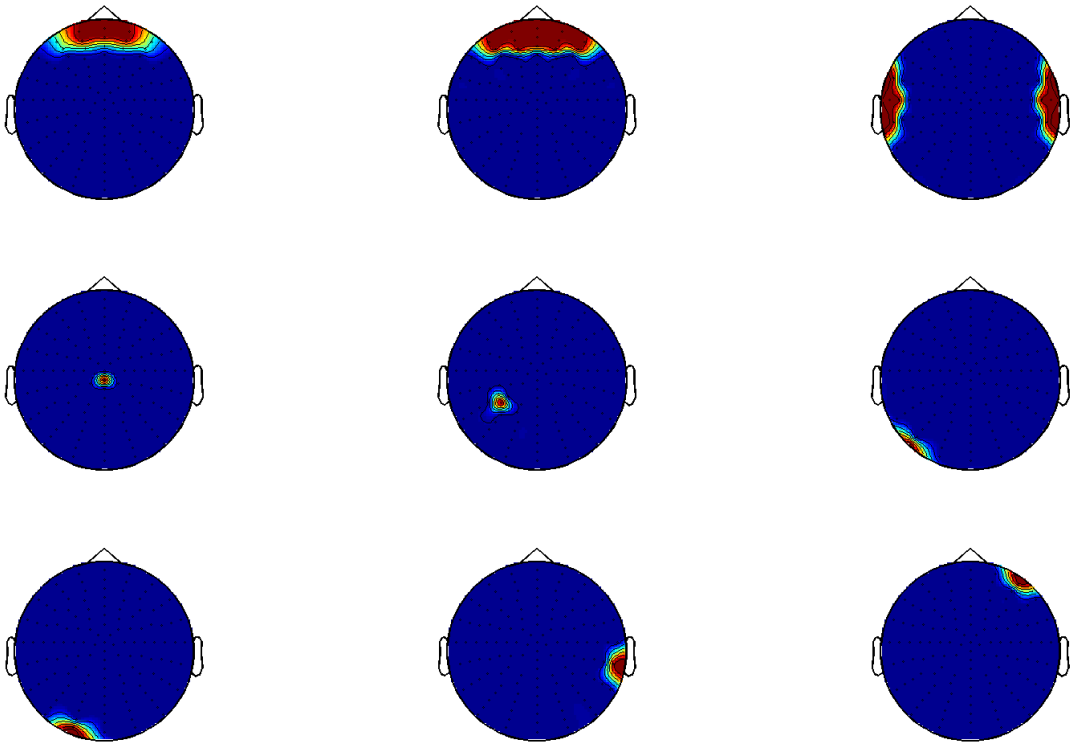


Figure 2.2 | Artifact Topography Template. Shown here are 9 examples of the template artifacts created for preprocessing. Top left and centre shows typical eye movement related activity. Top right shows typical artifacts arising due to the gap between ears and cap, or headphone related disturbance. The remaining examples demonstrate typical artifacts arising from noisy electrodes at various points across the scalp, and of varying degrees of severity. In each, blue = 0, and red = 1.

Why use ICA?

ICA has shown it to be a successful technique for identifying, separating, and removing artifacts in EEG data in a variety of experiments (Hoffmann & Falkenstein, 2008; Iriarte et al., 2003; Plöchl, Ossandón, & König, 2012). The consistent use and demonstration of successful cleaning makes it a reliable technique for data cleaning. ICA also has the potential to detect small or medium modulations in the EEG signal that may be related to artifacts (e.g. small eye blinks), which would be missed using an arbitrarily defined cut-off threshold value (which looks only for extreme minimum and maximum deviations). Finally, by removing components in the data rather than entire trials, you can (potentially) save more of your data, reducing the chance of wasting cost and time.

The main limitation when using ICA as a method of artifact removal is that it can be difficult (if not sometimes, impossible) to determine which components relate to true artifacts (e.g. noise, bad channels, and eye movements) and which relate to cognitive processes. Thus, one must be careful of how you identify and remove components decided to be artifacts.

In this thesis, determining which ICA components were artifacts was performed using three automatic procedures (described below) alongside visual inspection. Automatic procedures are beneficial as they allow you to identify artifact EEG components with some consistent standard between trials and across subjects. Additionally, automatic identification is faster than visual identification when dealing with large datasets. However, it is important to stress the need to perform both visual and automatic inspection; if you do not carefully check the components automatically identified as artifacts via visual inspection, you risk automatically rejecting valuable sections of the data, removing too many components, or even keeping data which contains artifacts that the automatic procedures have missed. Thus, a combination of automatic and visual inspection is essential.

Artifact identification via correlation with templates

As mentioned above, artifacts related to ocular-motor activity and noisy channels have distinctive topographies: eye movements and blinks show strong frontal activity, ear movement or headphone artifacts show strong activity centred over the ears, and noisy channels show strong activity centred over single electrodes. To detect and remove these artifacts, templates containing these typical topography patterns were manually created using Matlab for both the 64 channel and 128 Biosemi layouts. These were generated by inserting high amplitude values (i.e. +1 or 0.4 relative to 0) over several typical artifact areas: (a) centro-frontal regions (related to eye movements), (b) broader frontal regions (as in Hipp & Siegel, 2013), (c) temporal areas (related to temporal muscles, and headphones), and (d) single electrodes (related to noisy channels). Visual examples of these templates are shown in Figure 2.2, and the electrodes for which we created artifact values are listed in Table 2.1. Using these templates, the correlations between the signals in each ICA component and the signals in each topographical template were calculated and components with

64 Channel								
Frontal	Fp1	Fp2	Fpz	AF7	AF3	AFz	AF4	AF8
Broad Frontal	Fp1	Fp2	Fpz	AF7	AF3	AFz	AF4	AF8
Temporal	T7	T8	FT7	FT8	TP8	TP7		
128 Channel								
Frontal	C29	C17	C16	C28	C18	C15		
	C30	C31	C27	C19	C14	C9	C8	
Broad Frontal	C29	C17	C16	C28	C18	C15		
	C30	C31	C27	C19	C14	C9	C8	
Temporal	D8	D22	D23	D24	B27	B25	B26	B14

Table 2.1 | Artifact Template Electrode Labels. Table showing the electrodes for which we manually inserted artefactual amplitude values. These electrodes were chosen as they correspond to the locations typical of artifactual eye blinks and muscle movements around ears and eyes. Electrodes denoted in black text = 1 μV, electrodes denoted in underlined text = 0.4 μV). These values were chosen to create gradual fluctuations from baseline (zero) that show patterns typical of artifacts. In addition to the electrodes listed above, artefactual values of 1 were inserted over each individual electrode to generate a topographical template for each electrode representing a broken or noisy channel. For more detail, see section: Artifact identification via correlation with templates.

high correlations (defined in this case as >0.8) were suggested for removal. Visual inspection was used to confirm.

Artifact identification via correlation with electro-oculogram (EOG) signals

The second way artifacts were removed was via correlating the component time signals (identified by ICA) with electro-oculogram (EOG) signals. EOG signals capture the high frequency (>30Hz) EEG potentials that occur due to muscle movements in the face and neck, miniscule eye rotations, and involuntary microsaccades (Hipp & Siegel, 2013; Keren, Yuval-Greenberg, & Deouell, 2010; Muthukumaraswamy, 2013). Conventional filtering and ICA often miss these, and this can be particularly problematic if performing time-frequency analysis.

In this thesis, EOG signal channels (as defined in Keren et al., 2010) were calculated using the eye movement signals from four electrodes placed around

the eyes. These were: a vertical electro-oculogram (VEOG), a horizontal electro-oculogram (HEOG), and a radial electro-oculogram (REOG) signal (Croft & Barry, 2000; Elbert, Lutzenberger, Rockstroh, & Birbaumer, 1985; Keren et al., 2010). The VEOG picks up micro eyelid movements, and is calculated as the difference between signals in the electrodes placed above and below the eyes. The HEOG picks up small eye rotations (corneo-retinal dipole rotation), and is calculated as the difference between the electrodes placed at the outer corners of each eye. Finally, the REOG signal picks up saccade related spike potentials (SPs) (Boylan & Doig, 1988; Doig & Boylan, 1989; Thickbroom & Mastaglia, 1986), and is calculated as the difference of all averaged EOG channels placed around the eyes and a referenced posterior electrode (in this thesis, for 64 channel analysis the average of Pz, P1 and P2 were used, and for 128 channel analysis A17,18,20,21,30 and 31 were used). These three EOG signals were correlated with the signals derived from ICA, and any which had high correlations (again, defined as >0.8) were identified as artifacts. After visual confirmation, these were removed.

Artifact identification via power spectrum analysis

The final method used to remove artifacts was to compute the power spectrum of each ICA component, and remove components that had a low ratio between low frequency and high frequency power (with a low ratio defined as <6). In general, EEG data — or any data where there is a relationship between power and frequency — follows a power law c/f^x, where c is a constant and x is an exponent (Cohen, 2014). This law states that power (here, EEG power) is a function of frequency (here, EEG frequency). EEG data specifically follows a 1/f phenomenon, meaning that the signal has smaller magnitude at higher frequencies than the signal at the lower frequencies. By searching for components where there is a low ratio between low frequency power and high frequency power (i.e. little difference, which is in contrast to what we would expect), we can find components that are most likely related to artifacts and remove them. For this thesis, the power spectral density ratio of each ICA component was calculated using Welch's method (as implemented in MATLAB 2012, pwelch function), and components that had a ratio less than 6 were then checked again for correlations (this time, >0.5) between the component and the

artifact templates. These components were visually identified and removed if they appeared to be related to artifacts.

To summarise, in this thesis a combination of ICA, visual inspection and automatic procedures were used to remove artifact components in the EEG data. The following sections will describe all the analyses used on the cleaned EEG data in this thesis.

Time-Frequency Approach

As described in Chapter 1, the EEG activity measured from a brain is not static; signals evolve and fluctuate over time in rhythmic patterns. These rhythms — known as "neural oscillations" — are driven by excitatory and inhibitory fluctuations in signals from populations of neurons, and show variations in the speed with which they oscillate. This allows them to be subdivided into different bands: delta (δ, 1-4 Hz), theta (θ, 4-8 Hz), alpha (α, 8-12 Hz), beta (β, 13-30 Hz) and gamma (γ, >30Hz). Alongside temporal and spatial information, they contain information about frequency (the speed of the oscillation), power (the amount of energy) and phase (the point along the wave).

In neuroimaging, a time-frequency based approach tries to understand how these oscillatory patterns in the brain relate to different cognitive and perceptual processes. There are two important aspects to consider when generating or evaluating time-frequency representations (TFRs). The first is taper choice. A taper is a function multiplied with the EEG data, and is important for reducing spectral leakage (i.e. where the energy in a particular frequency band "leaks" to the corresponding bands) and controlling frequency smoothing (which is performed to remove localised variation in the raw EEG). In this thesis Hanning tapers were used. These multiply the time course data with an inverted cosine function, to generate a signal that peaks in the middle but tapers off to zero on either edge. The benefit of this taper is that it tapers the data fully to zero at the beginning and end of the time window, thus eliminating the possibility of edge artifacts. However, this means you lose the data at the edges, but using appropriate time windows will limit these effects. Shorter time windows provide better temporal precision but lower frequency resolution, while

longer time windows provide better frequency resolution but lower temporal resolution. Therefore, the choice of time window should depend on what frequencies you want to analyse and the trade-off you desire. As the goal of this thesis was to examine the temporal evolution of multisensory processes in the brain, we chose sliding time windows that scaled with specific frequency bands to minimise the trade-off: for lower frequencies there was a longer time window, while for higher frequencies the time window was shorter (the individual bands and time windows are listed in the Methods section of Chapter 4). This allowed us to keep temporal smoothing to a minimum with higher frequencies.

Finally, the time-frequency representations for the work in this thesis were generated using complex wavelet analysis. In short, this method examines how the frequency content of the signal changes over time by convolving the raw EEG data with a complex wavelet. It is a multi-step process which involves extracting an epoch of the data (based on the desired window length), Fourier transforming both the taper and the data, multiplying them together, and then computing the inverse Fourier transform. This analysis was repeated for each epoch in the trial, resulting in a power spectrum for the entire dataset.

Why use a Time-Frequency Approach?

The main benefit of using a time-frequency approach is that it generates a multidimensional representation containing information on time, space, power, frequency, and phase; this allows you to explore more than simply time and amplitude values in your signal. This is in contrast to the traditional method of studying the timing of neural responses: the event related potential (ERP). The ERP results from averaging EEG activity over trials to generate a one-dimensional time signal, which contains positive and negative fluctuations time-locked to a specific event (usually stimulus or response locked). The assumption behind this is that noise or any other signals not related to the cognitive process of interest will be averaged out, leaving only fluctuations of interest. However, averaging also removes a lot of the variation present in the signal. In contrast, the time-frequency approach allows you to examine the temporal, spatial and spectral properties of your EEG signal. As discussed above, the main limitation to the approach is the need to trade-off between temporal precision and frequency

specificity; if you want more temporal precision, you must examine higher frequency activity, and if you want to analyse low frequency activity you will have to give up temporal precision. Despite this limitation, the temporal resolution of time-frequency decomposed data is still on the order of milliseconds.

To summarise, the time-frequency approach is a useful way of investigating the temporal dynamics of EEG signals by examining the oscillatory nature of the signal and allowing you to compare differences between conditions or groups of subjects.

Linear Discriminant Analysis

An alternative way of examining the EEG signal is to use multivariate analysis methods. These techniques focus on extracting information across space and time from single trials, allowing you to examine inter-trial variability within subjects. This can reveal information that is hidden when you average across trials (see Figure 2.3), and allows you to relate subject specific behaviour to neural signals on a trial-by-trial basis (Blankertz, Lemm, Treder, Haufe, & Müller, 2011; Parra et al., 2002; Parra, Spence, Gerson, & Sajda, 2005; Pernet, Sajda, & Rousselet, 2011; Sajda, Philiastides, & Parra, 2009).

For the multivariate analysis in this thesis, a technique known as linear discriminant analysis (LDA) was used. LDA is a popular technique used for dimensionality reduction and data classification, which spatially integrates information across the electrodes rather than trials, thus providing a single-trial measure of cognitive function. It is becoming increasingly popular in neuroimaging, and has been used successfully to show information related to various stimuli and cognitive processes can be extracted from neural signals (Gherman & Philiastides, 2015; S. J. Kayser, McNair, & Kayser, 2016; Philiastides, Heekeren, & Sajda, 2014; Philiastides & Sajda, 2006).

The overall goal of LDA is to generate a one-dimensional projection (denoted as Y), which linearly combines the information from the multidimensional EEG data

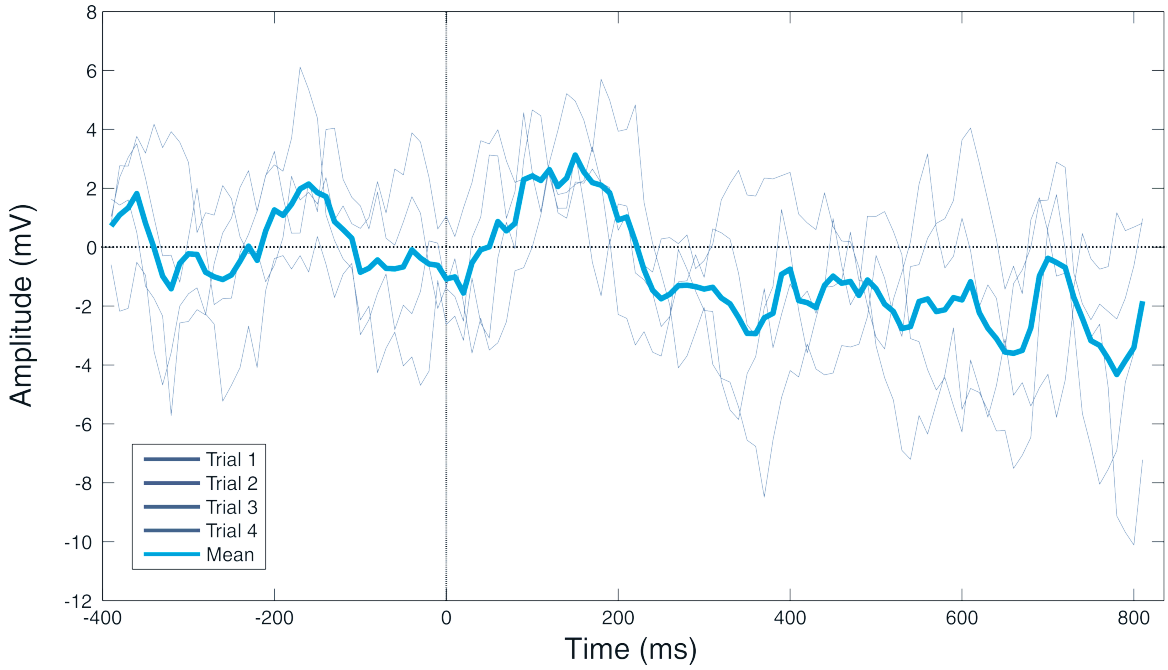


Figure 2.3 | Single-Trial Analysis. Schematic only. Grey lines show the individual trial signals. The blue line represents the trial averaged signal. As can be seen, there is variation amongst even very few trials, and this variation is lost due to averaging.

(X) into a single channel (*dimensionality reduction*). This projection (Y) is calculated from a set of spatial weights (denoted W), and represents the best linear separation of the EEG data into two "classes" based on predefined conditions of interest (*data classification*). Each step in this process will be described below.

Estimating the Spatial Weighting Vector (W)

The spatial weighting vector (W) is essentially a spatial filter (of the same dimension as your EEG electrode set) that represents the activity components most sensitive to the classification of the two conditions of interest. There are different methods available to calculate the weight vector, but in this thesis regularised fisher's linear discriminant (FLD) was used. FLD finds the best mean separation between the two conditions of interest, while minimising the overlap of the covariance matrices (see Figure 2.4 and Parra, Spence, Gerson, & Sajda, (2005) for tutorial). This is formalised with:

$$W = S_c (m_2 - m_1)$$

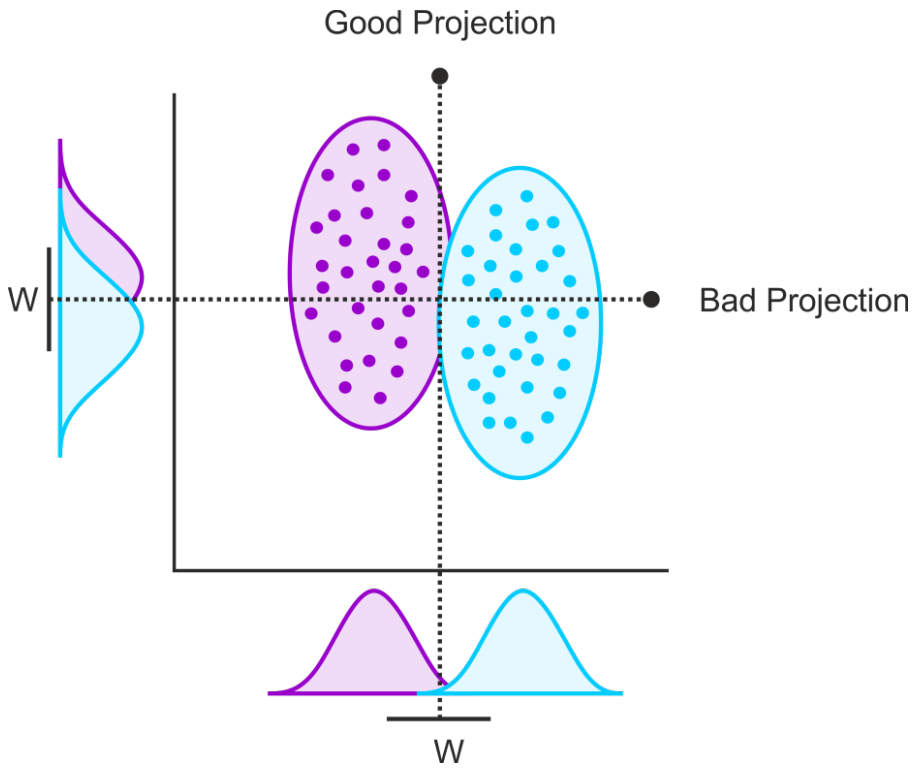


Figure 2.4 | Linear Discriminant Analysis (Fisher's). In the case of Fisher's linear discriminant, the goal is to find the projection (dotted lines) that best separates the given samples (coloured points) from two conditions of interest (here represented by, class 1 and class 2, blue and purple) while minimising the overlap of the covariance matrices (shaded circles). W represents the set of spatial weights used to find the projection (Y). The data presented here are a schematic only.

where m is the mean of condition 1 and 2 respectively (e.g. high tone (1) or low tone (2), or correct response (1) or incorrect response (2)), and S_c is the common covariance matrix (the average of the condition specific covariance matrices, see Philiastides, Heekeren, & Sajda, 2014). This covariance matrix is regularised to account for any potential estimation errors (e.g. overfitting) and the fact that the two conditions of interest will not have identical covariance structures (Philiastides, Heekeren, & Sajda, 2014).

Estimating the Projection (Y)

To estimate the discriminant projection (Y), the spatial weighting vector (W) is multiplied by the EEG activity alongside a constant at specific time points or time windows of interest:

$$Y_{(t)} = W_{(t)} X_{(t)} + C$$

The resulting projection (Y) is a one-dimensional aggregate representation of the EEG data at all sensors, which represents the best separation of the data (X) from two predefined conditions of interest. This projection is calculated for each trial and time point or time window of interest individually, resulting in a trial-specific, time-resolved measure of evidence about the conditions of interest. Additionally, the values assigned to the projection (Y) describe how informative the signal is over time: low Y values (i.e. those closer to zero) represent less information, while high Y values represent more information. Thus, the projection (Y) can essentially be used as a time sensitive, trial specific measure of condition evidence in neural signals.

Estimating the Forward Model (A)

To visualise the discriminant projection (Y) at the sensor level again, a scalp topography for each projection can be computed by estimating a forward model which linearly relates the sensor level EEG activity (X) to the discriminating component (Y), at each time point during the trial. This is formalised with:

$$A = X*Y \ / \ Y'*Y$$

where A is a spatial topography of the normalised correlation between the discriminant component (Y) and the EEG data (X) (and Y' represents the transpose of Y). Thus, A is a measure of the relationship between the EEG activity and discriminant projection, and can be used to show where this relationship is strongest (i.e. where there is the highest correlation between the projection and the data).

Classification Performance (Az)

The final part of the LDA process is to calculate how well the two conditions of interest can be separated. In this thesis this was carried out using the area under a receiver operator characteristic (ROC) curve combined with cross-validation. The ROC process calculates how sensitive and specific your classifier is (i.e. how many times the classifier correctly separates the conditions vs. how many times it incorrectly does it). Using it in combination with 10-fold cross-validation, involved randomly splitting the data into 10, equally sized partitions, training the classifier on 9 partitions, and testing the performance on the remaining partition. The average performance over these 10 partitions was then computed, resulting in one ROC value (Az) for each time point in the trial which representing how well the classifier performed. Finally, to assess whether this performance was above chance, permutation testing was used. This involved calculating the Az value (using the ROC analysis described above) for each time point 1000 times with randomly shuffled condition labels. This produced a probability distribution for Az, from which the percentiles were calculated, and the significance value that lead to a p-value of 0.01 (above the 99^{th} percentile) was extracted to represent chance performance level.

LDA: Final Outputs

Performing all of the steps above results in: a set of spatial weights (W) for each time point, a time-resolved discriminant output (Y) for each trial, a sensor level topography (A) which represents the strength of correlation between your discriminant output and EEG activity, and a measure of classifier performance (Az) for each time point in your trial. An example of each is shown in Figure 2.5.

Why use LDA?

There are three main benefits to using LDA. The first is that it preserves the trial-to-trial variability in the signal. This allows you to examine more of your data, find effects that may be lost due to averaging, and gives you the option to relate neural signals to behaviour on a single-trial basis. The second benefit concerns signal dimensionality. EEG data contains many dimensions (e.g. four

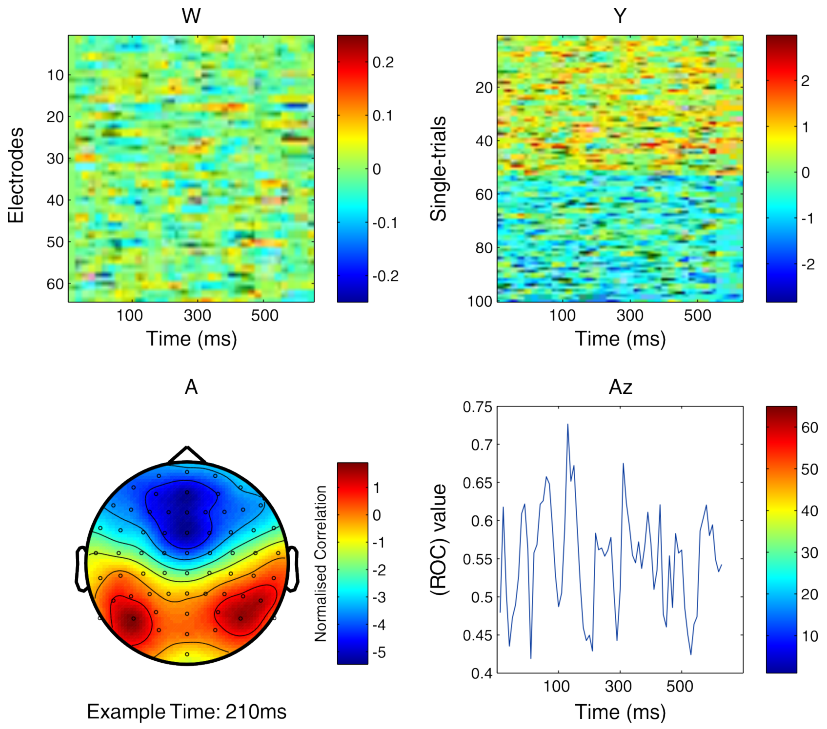


Figure 2.5 | LDA Output Example. Schematic of: the spatial weights (W) for one trial at each time point (*top left*), the discriminant output (Y) for individual trials sorted by condition type (*top right*) for each time point, an example of a forward model topography (A) calculated from the discriminant component and EEG data (*bottom left*), and the classification performance (Az) over time (*bottom right*) for one subject.

dimensions are typically: trials, electrode space, time points, and conditions). Data recorded with a 64-channel EEG system measuring at 1000Hz, with each trial lasting approximately 3 seconds, and approximately 2000 trials (as in Chapter 3), will result in 384,000,000 features for each subject (2000 trials x 64 electrodes x 3000 time points), which vastly outnumbers the number of observations they actually came from (i.e. one subject). Somewhat counter intuitively, as dimensionality increases the performance of your classifier decreases; as the feature space grows (with more dimensions) it becomes sparser, and it becomes easier to find a linear way to separate the separate the data. This can lead to overfitting, and the classifier will not generalise well when applied to new data. The dimensionality reduction aspect (concatenating over electrodes rather than trials) of LDA goes towards solving this. This leads to the third benefit, which is related to integrating the information across sensors rather than trials. Individual sensors are noisy (due to physiological and environmental noise), and often carrying similar information from neighbouring sensors. As the one dimensional projection is a representation of the activity from all sensors, it can be a better estimate of neurophysiological activity when

compared to a single electrode, thus increasing the signal-to-noise ratio of your signal (Blankertz et al., 2011; Parra et al., 2005).

There are also two limitations to consider. First is the assumption of linearity. Linear discriminant techniques assume that brain activity can be modelled linearly. While there are good reasons to consider that this may be true (e.g. linear models have been shown to predict neuronal response patterns in animal work ,Fetsch, Pouget, DeAngelis, & Angelaki,2012), there is also evidence that shows distinctly non-linear activity in the brain (Rombouts, Keunen, & Stam, 1995). Therefore, the assumption of linearity may not hold in all cases. Second, LDA assumes normally distributed data, independence of features (such as electrodes and time points), and identical covariance matrices for each condition you are classifying (e.g. trials belonging to one condition versus another), which is often not true in EEG data. Fortunately, regularising the covariance matrices can help control for effects of non-equal covariance structures (Blankertz et al., 2011; Philiastides et al., 2014), and some research has shown that LDA can be robust even if these assumptions of common covariance and normality are violated (Li, Zhu, & Ogihara, 2006).

To sum up, LDA is a powerful tool for dimensionality reduction and condition classification, and results in a time-resolved, single-trial measure of neural activity that specifically relates to functional differences between task relevant conditions.

Mutual Information Analysis

The final analysis technique used in this thesis was an information theoretic approach known as mutual information (MI) analysis. In short, MI measures the statistical dependency between two signals of interest (e.g. between stimulus type and EEG, or between EEG and reaction times), and allows for the direct comparison of the MI signals underlying different conditions (this will be explained in more detail below). Such analyses are becoming increasingly popular to compare different neuronal responses at the level of single neurons (Ince, Panzeri, & Kayser, 2013; Kayser, Montemurro, Logothetis, & Panzeri,

2009), and in neuroimaging (Gross et al., 2013; Ince, Jaworska, et al., 2016; Kayser, Ince, Gross, & Kayser, 2015; Keitel, Ince, Gross, & Kayser, 2017).

Calculating Mutual Information

The most basic element of MI is entropy, defined as the amount of information a variable has. Entropy is a measure of uncertainty, which can be used to examine how predictable a variable is. For example, take a random variable *X* that represents the card that you get when you draw a card from a pack. Then take a random variable *Y*, which represents the number that falls when you throw a six-sided die. Which has higher uncertainty (and thus, higher entropy)? In this example, the entropy of the card draw is higher: there are more possible options for a card draw (52 possible values) than a die throw (6 possible values). Now, imagine you are measuring the entropy of a single EEG electrode. Higher entropy means that the electrode you are measuring/comparing can take on a higher number of states, interpreted as the activity has more configurations over time. Lower entropy means the activity in the electrode is more restricted. As a result, you can see how knowing the entropy of a variable helps you understand how predictable or frequent it is. A related concept is conditional entropy, which is a measure of how much uncertainty remains about variable X, when you know the value of variable Y. For example, if we know how much one signal varies (Y), what does this tell us about another (X)? The conditional entropy (X|Y) of the two random variables X and Y can tell us this.

Finally, mutual information is defined in terms of entropy differences; it quantifies the reduction in uncertainty (entropy) about one variable (X) given the knowledge about another variable (Y). As such, MI measures the entropy explained about X given Y (for a visual representation, see Figure 2.6). In this case, high MI between two signals indicates a large amount of explained entropy (i.e. reduction in uncertainty) about Y given X, and low MI indicates a small amount of explained entropy (or, small uncertainty reduction). Zero MI indicates that the two variables are independent. However, MI is an unsigned quality (i.e. never below zero), and so can reveal only the strength of a relationship rather than the direction.

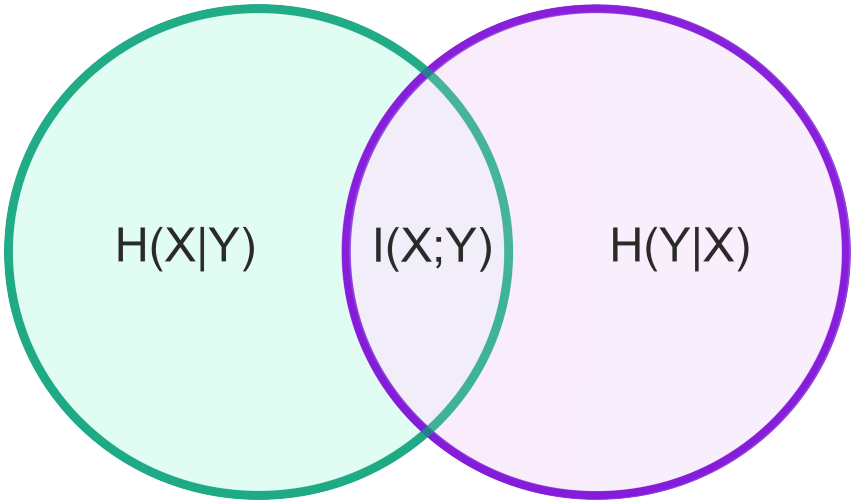


Figure 2.6 | Mutual Information. Schematic shows how mutual information (I) is a measure of the shared information (explained entropy) shared by X and Y.

In this thesis, a novel method known as Gaussian Copula Mutual Information (GCMI) was used. GCMI uses the concept of statistical copula and Gaussian transformation to calculate MI. In short, the distribution of each random variable (in our case, the EEG signal, and the behavioural measure) is transformed to be normal via statistical copulas, and then MI is estimated by applying a Gaussian parametric estimate (see Ince et al., (2016) for a full tutorial). This method is computationally efficient and can be applied easily and robustly to multidimensional EEG data using the framework devised by Ince et al., (2016) and the code is freely provided in the GCMI toolbox (see Ince et al., (2016) for code and tutorials).

Why use MI analysis?

First, as MI is a measure of the uncertainty explained about X given Y, it can be used as a measure of the information that X and Y share. Consequently, if you find high MI then you can make the strong inference that the X (e.g. your stimulus) explains a lot about Y (e.g. your signal). Second, because MI quantifies the difference in entropies (which is a mathematically defined construct), it can be used a test for statistical dependence between two variables (see Ince et al., 2016). Third, it does not have any assumptions about the underlying distributions or the nature of the relationship between variables, and is therefore more

flexible than other statistical analysis methods. Fourth, MI is a general, quantifiable measure quantified on a common scale ("bits"). This common scale allows for the direct comparison of MI values between signals from a range of measures: you can compare the MI values resulting from behavioural and EEG signals, from different experiments in the field, and from studies in the field using different techniques (e.g. between single-neurons and neuroimaging signals). Finally, the MI measurement is sensitive to nonlinear effects (Ince, Giordano, et al., 2016; Ostwald & Bagshaw, 2011), which analysis methods such as linear discriminant analysis are not. This is a benefit as nonlinearities may play a role in the neural signal.

To summarise, MI is a simple but robust method for quantifying the amount of information that is shared between two variables, and is essentially a flexible, statistical test for significant dependence between two signals.

Chapter 2: Summary

There are various methods used in EEG research today that allow us to explore our data and answer our questions. These all have their benefits and costs, as have been discussed in this Chapter. This overview should allow the reader to evaluate the experimental work in the following chapters in enough detail to conclude the suitability of the analysis given the specific experimental paradigms, and to interpret the results in the appropriate ways.

Chapter 3 : Neural Correlates of Multisensory Reliability and Perceptual Weights Emerge at Early Latencies during Audio-visual Integration

Introduction

The reliability of the information received by our senses varies. For example, visual cues become unreliable in dim or fogged conditions, and auditory cues become unreliable in loud or noisy situations. To deal with such variations, stimulus reliability must be taken into account when combining information across the senses. “Optimal Integration” or “Bayesian Integration” models formally describe how this process occurs. These models state that observers can arrive at the most precise final perceptual estimate by weighting sensory information in proportion to its reliability; more reliable cues are assigned a higher weight and have stronger influence on the final perceptual estimate, while less reliable cues are assigned a lower weight and have less influence. In most cases, this weighting leads to a more precise and reliable percept (Angelaki et al., 2009; Ernst, 2006; Ernst & Bülthoff, 2004; Fetsch, DeAngelis, & Angelaki, 2013b; Rohde et al., 2015). Formally, these models have two predictions: first, that performance should be higher (or perceptual thresholds should be lower) for multisensory trials compared to unisensory trials and second, as reliability of the cue decreases, so does the importance (or weight) the subject assigns to it.

These models have had great success in demonstrating how integration occurs in wide variety of tasks a wide variety of behavioural tasks. For example, Raposo et al., (2012) and Sheppard et al., (2013) examined how subjects (rats and humans) weighted auditory and visual as the reliability of both cues changed. They used an audio-visual rate discrimination task that contained visual only, auditory only, and bimodal (audio-visual) trials, and where the reliability (high/low) and congruency between the auditory and visual cues was systematically varied. Their results revealed two outcomes. First, they found that the performance of rats and humans improved on multisensory trials compared to unisensory trials, and that these performances were close to the optimal predictions. Furthermore, they found that at a perceptual level humans’ and rats’ decisions about rate were influenced by the reliability of the auditory and visual stimuli.

Specifically, when the reliability of the two cues were matched, subjects weighted both modalities equally, but when the reliabilities were unequal, subjects consistently assigned higher weight to the more reliable cue. These observed cue weightings were also consistent with the theoretical weights predicted by statistically optimal models. Another study by Alais and Burr, (2004) investigated the integration of audio-visual cues during a spatial localisation task. Light spots (visual) and clicks (auditory) were presented to participants separately (unimodal condition) or together (bimodal condition), and participants had to indicate which of the stimuli were localised further to the left. They also manipulated the reliability of the visual information by blurring the light spots, and introduced incongruencies between the locations of the stimuli. Their results showed that in all subjects, localisation in the bimodal conditions was better than localisation in either the visual alone or auditory alone condition, consistent with the first key prediction of the optimal observer models. Additionally, they found that when the visual information was reliable (i.e. not blurred), vision dominated localisation, but when the visual stimuli were not reliable (i.e. blurred) the auditory information dominated. Again, this is consistent with a simple model of optimal combination of the auditory and visual information, where the two are weighted linearly in proportion to their reliability. These and numerous other studies have shown that cue integration follows the rules of optimal integration for audio-visual (Alais & Burr, 2004; Battaglia et al., 2003; Rohe & Noppeney, 2016), visual-tactile (Beauchamp et al., 2010; Ernst & Banks, 2002; Helbig et al., 2012; Helbig & Ernst, 2007), and visual-vestibular stimuli (Butler et al., 2010; Fetsch et al., 2012, 2009; Morgan et al., 2008).

Weighted linear cue combination can also be encoded in neural signals. For example, Fetsch et al., (2012) recorded neurons in the medial superior temporal dorsal area (MSTd) of macaques while they were performing a visual-vestibular heading discrimination task where the reliability of both modalities varied. By fitting neurometric curves (a psychometric curve fit to neural responses) to the electrophysiological data, they were able to show that both behaviour and neurons responded in line with the rules of optimal integration: specifically, neurons responded with an increase in firing for the more reliable cue. In another study, Morgan et al., (2008) demonstrated that the responses from

neurons in the MSTd stimulated with visual-vestibular stimuli were fit well with a linear weighted model, and that the weights changed according to the relative reliabilities of the stimuli. Again, as visual reliability decreased, the weight needed to fit the model increased. Thus, animal work suggests that probability based encoding can be implemented in neural circuits, and that the principles of statistically optimal integration can be encoded in the firing patterns (at least, in area MSTd).

Neuroimaging data can also be combined with statistically optimal frameworks to study cue integration. For example, Helbig et al., (2012) used fMRI to characterise the how neural signals were modulated by changes in visual reliability during a visual-tactile integration task. Their results showed a significant increase in the blood oxygen level dependent (BOLD) response in the left and right postcentral sulcus and left superior parietal lobe when the reliability of the visual stimulus was reduced. They also found decreased activation in the right posterior fusiform when the visual input was completely blurred. Together these results indicated that cue reliability seems to be reflected in population level BOLD activity, with increased activation in relevant sensory areas elicited by reliable stimuli, and suppression of activity in relevant areas by unreliable stimuli. Another fMRI study by Beauchamp et al., (2010) modelled the connection weights between occipital, somatosensory and parietal cortex during a visual-tactile task while manipulating the reliability of the stimuli. Using the average BOLD signal in pre-defined regions of interest (ROIs), structural equation models were constructed which reflected the correlation of activity between these ROIs (for each ROI pairing and subject separately). Their results showed that the connection weight between somatosensory and the intraparietal sulcus (IPS) increased when the tactile stimulus was more reliable, yet the connection weight between occipital cortex and IPS increased when the visual stimulus was more reliable. Thus, their results not only show cue weighting emerging in the brain, but also suggest it may be implemented via connections and cross-talk between early sensory regions and parietal cortex. Finally, Rohe & Noppeney, (2016) examined where in the cortex audio-visual signals were weighted using a spatial ventriloquist paradigm where the reliability of the visual stimulus varied. Their results also showed linear cue weighting evident in the BOLD signal, but only in parietal cortex; there were no effects of

sensory weighting at low-level areas. Overall, these results indicate that linear cue combination can be implemented by neurons in the brain, and that these computations can be measured by neuroimaging.

However, while providing valuable computational insights, these studies have some limitations. Single-neuron recording studies cannot adequately assess the spatial distribution of cue weighting processes across the cortex, and fMRI studies cannot determine the temporal evolution of the neural processes implementing sensory cue weighting. Therefore it remains unclear when following stimulus onset neuronal responses across the cortex are modulated by the reliability of sensory information, and when they reflect the modulations in subsequent perceptual weighting.

To address these questions, in this study we combined a rate discrimination task with EEG based neuroimaging, single-trial decoding, and linear modelling. Importantly, this combination allowed us to examine the temporal dynamics of weighted cue combination, to identify the neural correlates relating to sensory reliability, and to link modulations in neural signals to perceptual reweighting processes on a trial-by-trial and time-resolved basis.

Methods

Subjects

We obtained data from 20 right-handed subjects (13 females; mean age 26 years) after written informed consent. The sample size was set to 20 based on sample sizes used in related previous EEG studies. All subjects reported normal or corrected to normal vision, normal hearing, and received £6 per hour for their participation. The study was approved by the local ethics committee (College of Science and Engineering, University of Glasgow) and conducted in accordance with the Declaration of Helsinki.

Stimuli and Task

The task was an adapted version of a 2-alternative forced choice rate

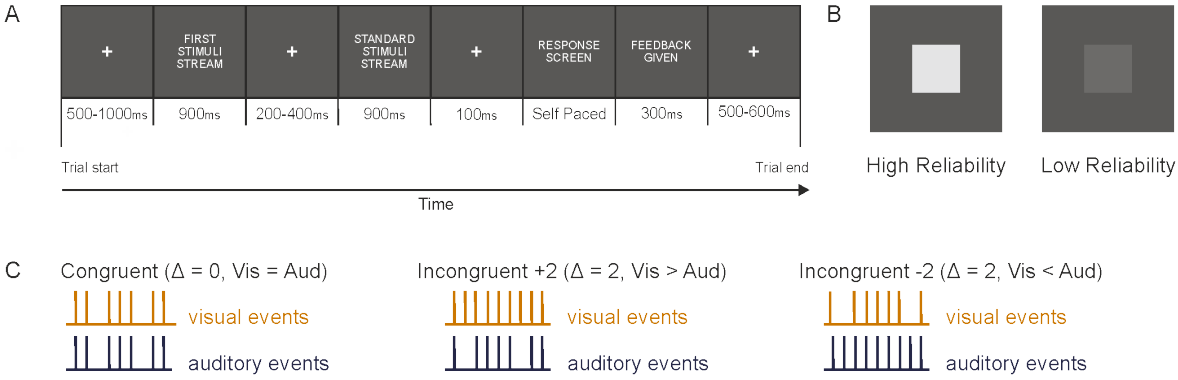


Figure 3.1 | Experimental Task. (A) Subjects were presented with two sequential streams of auditory, visual and/or audio-visual events (one experimental, and one comparison of 11 Hz) and had to indicate which contained more events. The first stream varied in modality, event rate, reliability and congruency of the events (see Methods). (B) Example of high and low reliability visual stimuli. (C) Schematic showing one combination for each level of congruency (left: equal rates, middle: auditory fewer events, right: auditory more events). Δ = Visual minus Auditory rate.

discrimination task (Raposo, Sheppard, Schrater, & Churchland, 2012; Sheppard et al., 2013). Subjects were presented with two streams (each lasting 900 ms) of auditory, visual or audio-visual events and asked to decide which stream had a higher event rate (Figure 3.1A). Visual events were noise squares (3x3cm, 2.1° of visual angle, flashed for 12 ms each) presented atop a static pink-noise background image. Acoustic events were brief click sounds (65 dB SPL, 12 ms duration) presented in silence. These events were instantiated by sequences of short (48 ms) or long (96 ms) silent pauses, causing them to appear as streams of auditory and/or visual flicker. Auditory stimuli were presented using Sennheiser headphones and visual stimuli were presented on a Hansol 2100A CRT monitor at a refresh rate of 85 Hz. All recordings were carried out in a dark and electrical shielded room. The experiment was controlled through MATLAB (MathWorks) using the Psychophysics Toolbox Extensions (Brainard, 1997) and custom written scripts.

In the first "experimental" stream, events were presented at seven different rates (8 – 14 Hz). In the second "standard" stream, events flickered at 11Hz. The total stream duration (900 ms) was predetermined based on the refresh rate of the computer (85 Hz) and desired event length (12 ms). From these parameters we calculated the possible rates that could be presented alongside the short and

long pauses while staying within the 900 ms time window (8 to 14 Hz). For each trial, short and long pauses were randomly placed around the events to create stimuli streams of equal length for all rates. For the comparison stream (where the event rate was always 11 Hz), pauses were randomly interspersed around 11 events on each trial.

Important to note, that on some trials the experimental stream rate was equal to the comparison stream rate. Subjects were unaware of this, but were still asked to make a choice about which stream had a higher event rate. In theory, this could result in a bias towards inequality; for example, if on 11 Hz trials, subjects always chose to respond that the rate was lower, this would create a bias towards more "first stream lower" responses. To control for this potential systematic influence on behaviour, we ensured these trials only made up a small percentage of each block (2% of trials per condition), and were randomly intermixed within each block.

Contrast levels for each reliability level were derived to match individual subject's psychometric thresholds in separate calibration blocks carried out prior to the main experiment (see Procedure). The reliability of the visual stimulus was manipulated by adjusting the contrast of the visual square relative to the background (Figure 3.1B). Auditory reliability was held constant throughout. Manipulating the reliability of only one modality allowed us to keep the experiment at a reasonable length (~3 hours per session) while accommodating the additional time necessary for EEG set up and extended inter-trial intervals (to include a baseline period). This manipulation of only a single modality's reliability is also in line with past work (Fetsch et al., 2012; Helbig et al., 2012; Rohe & Noppeney, 2016).

Congruency was manipulated by introducing differences in the event rate between the auditory and visual streams; as mentioned in the Introduction, placing the audio-visual cues in conflict is necessary to allow the calculation of perceptual weights based on subject's bias towards each sensory cue. Audio-visual trials were either Congruent ($\Delta = 0$) with auditory and visual streams each having the same number of events, or incongruent, with the visual either containing two more ($\Delta = +2$), or two fewer ($\Delta = -2$) events than the auditory stream (Figure 3.1C).

Overall these stimulus manipulations resulted in: three unisensory conditions (auditory [AUD], visual high [VH] and visual low [VL]); two congruent multisensory conditions (one where both the streams were highly reliable [AVH Δ = 0], and one where the auditory had high and the visual low reliability [AVL Δ = 0]); and four incongruent audio-visual conditions (AVH Δ = +2, AVH Δ = -2, AVL Δ = +2 and AVL Δ =-2). Within each audio-visual condition (AVH, AVL), there were auditory (e.g. AVH auditory) and visual (e.g. AVH visual) stimuli.

Experimental Procedure

Each session started with two unisensory calibration blocks, used to set the reliability levels of the auditory and visual stimuli (see, Stimuli). The stimuli in these calibration blocks were the same stimuli used in the experimental blocks, however only the easiest comparison rates (8 Hz and 14 Hz) were presented. The auditory calibration block consisted of 30 trials with the auditory stimuli presented in silence (highly reliable auditory stimuli). For auditory trials an overall performance score was calculated. The visual calibration block consisted of 150 trials (30 trials x 5 signal-to-noise ratios, SNRs), where the reliability of the visual stimulus varied systematically from high to low. For the visual trials psychometric functions were fit to the data, and two SNR levels for visual reliability selected from the resulting psychometric curve. Visual high reliability was set as the SNR at which visual performance was equal to performance on the auditory calibration block. Visual low reliability was set at the SNR at which performance was ~30% lower than auditory performance. These reliability levels were set at the start of each experimental session and held constant throughout.

Each experimental block began with a reminder of the task instructions. Each trial began with a white fixation cross presented centrally on a dark grey noise image (500-1000 ms). This was followed by the first “experimental” stimuli stream (900 ms), a fixation period (200-400 ms), and then the “standard” stimuli stream (900 ms). The experimental stream varied in the modality, rate, and reliability of the stimulus, on each trial and within each block. The standard stream was matched to the first experimental stream with regards to reliability and modality. Congruency (always Δ = 0) and rate (always 11 Hz) were not manipulated in the standard stream. At the end of the trial, subjects were cued to respond using the left (“first stream has more events”) or right (“second

stream has more events") keyboard buttons, and received feedback on their performance (Figure 3.1A). For audio-visual trials where the rate of the auditory and visual stimuli differed (i.e. $\Delta = \pm 2$), feedback was generated using the average rate of the auditory and visual streams. For trials where the rates in the experimental and standard stream were equal, feedback was randomly generated. Although providing feedback is not standard practice for studies investigating reliability weighting, previous work (Raposo et al., 2012; Sheppard et al., 2013) has shown no difference in performance for subjects who received feedback compared to those who did not. Therefore we chose to provide feedback in order to engage subjects with the task over the long experiment. Overall, each block consisted of 510 trials with modality (auditory, visual, audio-visual), reliability (visual high and low), event rate (8-14Hz) and congruency (audio-visual $\Delta = 0, \pm 2$) varying pseudo-randomly across trials in the experimental stream (see Stimuli and Task).

In total, subjects completed 2040 trials, split up into four individual blocks. These blocks were carried out across multiple sessions over two, three or four days depending on the subjects' availability and level of alertness during the session. Importantly, before the main analysis we checked for confounding learning effects which may have arisen due to these multiple sessions. The results showed there was no difference in performance or perceptual weighting strategies across blocks or days (see Supplementary Figure S3.1 A and 3.1 B for individual day performance and weighting).

EEG Recording and Preprocessing

EEG data were recorded using a 64-channel BioSemi system and ActiView recording software (Biosemi, Amsterdam, Netherlands). Signals were digitised at 512 Hz and band-pass filtered online between 0.16 and 100 Hz. Signals originating from ocular muscles were recorded from four additional electrodes placed below and at the outer canthi of each eye.

Data from individual subject blocks were preprocessed separately in MATLAB using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom written scripts. Epochs around the first stimuli stream (-1 to 2s relative to stream onset) were extracted and filtered between 0.5 and 90 Hz

(Butterworth filter) and down-sampled to 200 Hz. Potential signal artefacts were removed using independent component analysis (ICA) as implemented in the FieldTrip toolbox (Oostenveld et al., 2011), and components related to typical eye blink activity or noisy electrode channels were removed (see Chapter 2, Preprocessing section). Horizontal, vertical, and radial EOG signals were computed using established procedures (Hipp & Siegel, 2013; Keren, Yuval-Greenberg, & Deouell, 2010 and see Chapter 2, Preprocessing section), and trials during which there was a high correlation between eye movements and components in the EEG data were removed. Finally, remaining trials with amplitudes exceeding ±120 μV were removed. Successful cleaning was verified by visual inspection of single trials. For one subject (S20), three noisy channels (FT7, P9, TP8) were interpolated using the channel repair function as implemented in the FieldTrip toolbox.

Analysis Methods

Psychometric performance and Optimal Integration Model

For each subject, modality, and stimulation rate, the proportion of "first stream had a higher event rate" responses were calculated and cumulative Gaussian functions fit to the data using the psignifit toolbox for Matlab (Fruend, Haenel, & Wichmann, 2011). The threshold (s.d., σ) and the point of subjective equality (PSE, μ) were obtained from the best fitting function (2000 simulations via bootstrapping). These measures were then used to calculate a set of predicted and observed perceptual weights (Fetsch et al., 2012) for each modality in each audio-visual reliability condition. Predicted weights reflect the weights that an optimal observer would assign to each sensory cue in multisensory conditions (Fetsch et al., 2012). These were calculated using the thresholds (σ) from the PSE curves calculated from unisensory trials using:

$$W_{AUD} = \frac{\frac{1}{\sigma^2_{AUD}}}{\frac{1}{\sigma^2_{AUD}} + \frac{1}{\sigma^2_{VIS}}}$$

Equation 1

Observed perceptual weights represent the apparent weight a subject assigns to each sensory cue. These were calculated from the PSE (μ) from multisensory trials using:

$$W_{AUD} = \frac{\mu_{AV(\Delta)} - \mu_{AV(\Delta==0)} + \frac{\Delta}{2}}{\Delta}$$

Equation 2

where Δ represents the incongruency between the auditory and visual stimuli (Fetsch et al., 2012). For both perceptual and observed weights we assumed that auditory and visual weights sum to one:

$$W_{VIS} = 1 - W_{AUD}$$

Predicted and observed weights were derived for each modality and reliability separately, and averaged over congruency levels (Δ = ± 2).

Time-dependent Perceptual Weights

To examine how perceptual weights evolved over the course of a trial we modelled the relationship between sensory evidence and behavioural reports at each time point. As a measure of sensory evidence, we used the "accumulated rate" (see Figure 3.2), defined as the average number of stimulus events presented up to each time point in the trial (with an event defined as an auditory click, a visual flash, or audio-visual clicks and flashes presented together). Accumulated rate was calculated in increasing 12 ms time bins (as our stimuli were each presented for 12 ms) which resulted in 75 time points for the full stimulus stream of 900 ms. This yielded a trial and time-specific measure ("accumulated rate") of the experienced sensory evidence.

This measure of sensory evidence was then regressed against behavioural choice (first stream higher vs. second stream higher) for each trial, modality within each audio-visual condition (AVH auditory and visual, AVL auditory and visual), and time point separately using logistic regression (see Figure 3.2).

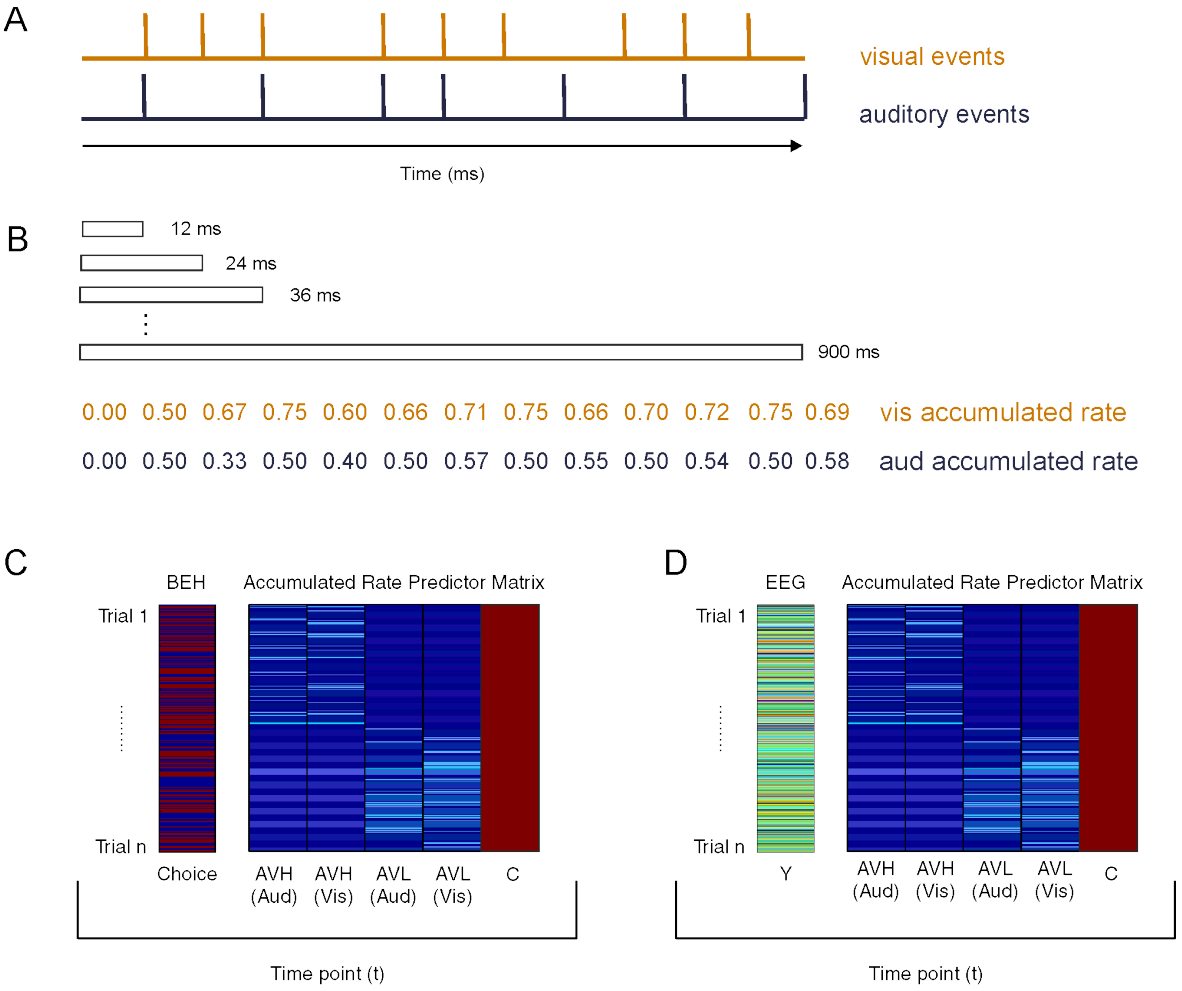


Figure 3.2 | Schematic of Accumulated Rate and Regression (Logistic and Linear) Analysis Pipeline (A) Step 1: the stimuli (auditory and visual) presented on each audio-visual trial are extracted. (B) Step 2: at each time window (increasing 12 ms time windows) the "accumulated rate" is calculated by taking the mean number of stimuli presented over the time window. For example, for the first two time windows no stimuli had been presented and so both the visual and auditory accumulated rates at these time points are zero. For the third time window, one visual stimulus had been presented and one auditory, so at this time point both accumulated rates equal 1/3 = 0.33 (1 stimulus/3 time points). For the fourth time window, two visual stimuli had been presented and still only one auditory stimulus had been presented; now the visual accumulated rate is 2/4 = 0.50, and the auditory is 1/4 = 0.25. This process is carried out for audio-visual high (both auditory and visual are of high reliability) and audio-visual low (auditory is high reliability, visual is low reliability) trials separately, and continues until you have a time varying accumulated rate for each modality, for each audio-visual trial. Finally, at each time point, the accumulated rates for both reliabilities are combined into one predictor matrix (with four separate predictors, one for each modality in each reliability condition) with a constant (see C, D). (C) Step 3: to calculate perceptual weights, logistic regression is used to regress the predictor matrix of accumulated rates generated in Step 2 against behavioural choice on a trial by trial basis, at each time point of the trial. This yields four regression weights (here defined as "perceptual weights") for each modality in each reliability condition for each time point in the trial. (D) Step 4: to calculate neural weights, linear regression is used to regress the predictor matrix (accumulated rates, same as above) against the decoding signal (Y) at each time point of the trial. Again, this yields four regression weights (here defined as "neural weights") for each modality in each reliability condition, for each time point in the trial. Note: in (C) and (D), for this early time point the accumulated rate is 0 for some trials in the predictor matrices. Data are schematic only.

Analysis was restricted to incongruent audio-visual trials and a time window of 24 ms to 600 ms post stimulus onset to account for null values (pre-24 ms) and multicollinearity in the predictor matrix (post 600 ms). This generated four sets of regression weights (which we here refer to as "perceptual weights"): one for the auditory and visual in each reliability condition (AVH auditory, AVH visual, AVL auditory and AVL visual).

To assess whether the accumulated rate was significantly predictive of perceptual choice, we quantified the performance of the regression model (referred to here as Az) using the area under the receiver operator characteristic (ROC) and 10-fold cross-validation (see Statistics). To determine how well the perceptual weights derived from the psychometric curves corresponded to the perceptual weights derived from the regression model, the correlation between the reliability influence for the psychometric and the reliability influence for the regression weights was computed at each time point during the trial (see Statistics). Here, the reliability influence was here defined as the difference (D) in auditory and visual weights at each time point (t) (in other words, the effect of visual reliability on auditory weights):

$$\mathrm{D(t)} = [\mathrm{AVH_{WAUD}} - \mathrm{AVH_{WVIS}}] - [\mathrm{AVL_{WAUD}} - \mathrm{AVL_{WVIS}}]$$

Equation 3

Single-Trial EEG Analysis

We used single-trial, multivariate linear discriminant analysis (Parra et al., 2005; Philiastides et al., 2014, and *see* Chapter 2, Methods) to uncover EEG components that best separated our two conditions of interest, defined here as stimulus rate higher (condition 1) or lower (condition 2) than 11 Hz. This analysis generated a one-dimensional projection (Y_t) of the multidimensional EEG data (X_t), defined by spatial weights (W_t) and a constant (C):

$$\mathrm{Y_t} = \mathrm{W_t}\,\mathrm{X_t} + \mathrm{C}$$

Equation 4

where the weight vector (W) represents the activity components most sensitive to the condition of interest (here, whether stimulus rate >/< 11 Hz), and the discriminant output (Y) provides a neural correlate of the quality of the single-trial evidence about the conditions of interest. This approach preserves the trial-to-trial variability of the data, and is assumed to be a better estimator of the underlying single trial task-relevant activity than the data on individual channels (Blankertz, Lemm, Treder, Haufe, & Müller, 2011; Kayser et al., 2016; Parra et al., 2005; Philiastides et al., 2014).

The classifier was trained to discriminate between high and low stimulus rate (i.e. whether the first stream had an event rate that was lower or higher than the standard stream of 11 Hz) as this reflected the task the subjects were asked to complete. The classifier was based on regularised (Fisher's) linear discriminant analysis (Philiastides et al., 2014), and applied to the EEG activity at each 5 ms time point from stimulus onset to 600 ms post stimulus onset in sliding time windows of 55 ms. For each time point, the EEG data within the 55 ms window was averaged and the discriminant output (Y) aligned to the onset of the 55 ms window. This time window was chosen as between each stimulus (of 12ms) there was either a 48 ms or 96 ms pause. Therefore, this window did not blur across stimuli and was in keeping with time windows used in previous work (Philiastides et al., 2014).

This analysis generated a sensory matrix [trials x time], which represented the information about stimulus rate in neural signals over time. To avoid introducing bias to either sensory modality, we derived the weighting vector (W) and constant (C) from the congruent audio-visual trials only (AVH and AVL $\Delta = 0$), and applied these to all other trials at the same time point. Scalp topographies for the discriminating component were estimated via the forward model (Philiastides et al., 2014), defined as the normalised correlation between the discriminant output and the EEG activity.

Neural Weights

We used linear regression to calculate a set of neural weights which quantified how each modality contributed to the information contained in the discriminant output (Y). Similar to the behavioural data, the trial-specific accumulated rates

for each modality, in each reliability condition (AVH auditory, AVH visual, AVL auditory, and AVL visual) were used as predictors and regressed (in a single regression) against the discriminant output (Y) at each time point in the trial (see Figure 3.2). Analysis was performed over the significant classification performance time window (24 ms to 400 ms), and restricted to incongruent audio-visual conditions ($\Delta = \pm 2$).

Neural weights were generated for each modality and reliability condition separately; this resulted in four neural weights for each time point (one for AVH auditory, AVH visual, AVL auditory and AVL visual). Importantly, the precise neural origin of these EEG discriminant output components (and hence the respective generators of auditory and visual contributions to these) remains unclear, and so we did not assume that auditory and visual neural weights for each reliability normalised to a fixed sum of one. For this reason, we did not perform normalisation on the neural regression weights. Finally, to assess the relationship between neural weights and the time-dependent perceptual weights (calculated using regression analysis), we correlated the reliability influence (Equation 3) between these two measures at each time point (see Statistics).

Source Localisation

We used source localisation to obtain a confirmatory representation of the neural generators underlying the discriminant output (Y). This involved correlating the source signals with the discriminant output (Y), and is comparable to obtaining forward scalp models from linear discriminant analysis (Parra et al., 2005).

The standard Montreal Neurological Institute (MNI) magnetic resonance imaging (MRI) template was used as the head model and interpolated with the automated anatomical labelling (AAL) atlas. Leadfield computation was based on the standard source model (3D grid model with 6mm spacing) and a manually aligned BioSemi electrode channel template. Covariance matrices were calculated from -200 ms pre-stimulus onset to 800 ms post-stimulus onset and source localisation performed on individual subject single-trial data using a linear constrained minimum variance beamformer in Fieldtrip (fixed orientation, 7% normalisation). This resulted in 11432 unique grid points within the brain for

which the source signal was extracted. Finally, source signals were correlated with the discriminant output (Y) for each subject and time point within our analysis window (24 ms to 400 ms) separately, and correlation signals z-transformed and averaged over subjects.

Statistics

All descriptive statistics reported represent median values. All Z values reported were generated from a one-sided Wilcoxon signed rank test (after testing assumptions of normality, which did not hold), and effect sizes calculated by dividing the Z value by the square root of N (where N = the number of observations rather than subjects, Rosenthal, 1994). Correlations were calculated using Spearman rank correlation analysis. All reported p-values were checked for inconsistencies using the R software package "statcheck" (Nuijten, Hartgerink, Assen, Epskamp, & Wicherts, 2016).

Significance levels of classification performance (Az) were determined by randomly shuffling the data by condition 2000 times, computing the group averaged Az value for each randomisation, and taking the maximal Az value over time. This generated a distribution of group averaged Az values based on 2000 randomised data sets, from which we could estimate the Az value leading to a significance level of $p<0.01$. Significant clusters for all other comparisons were determined using a cluster based randomisation technique (referred to as cluster randomisation in text, Maris & Oostenveld, 2007). In all cases, condition labels were randomly shuffled across conditions 1000 times, and for each separate comparison a distribution of t-values was computed. These shuffled t-value distributions were compared to the true t-value values to find significant clusters in the data: for all comparisons a cluster-threshold of $t = 1.8$ (corresponding to $p<0.01$), minimum cluster size of 2, and max-size as the cluster-forming variable were used. Effect sizes were calculated as the equivalent r value that is bounded between 0 and 1 (Rosenthal & Rubin, 2003).

To improve readability in the following results section, results are reported with regards to significance thresholds (e.g. <0.05) within text, while exact p-values and statistics are reported in corresponding tables.

Results

Psychometric behaviour and perceptual thresholds

Figure 3.3A shows the group-level psychometric curves for each sensory condition. Unisensory curves (Figure 3.3A, left) demonstrated that subject performance was lowest for visual low reliability conditions and highest for auditory, as evidenced by steepness of curve. Comparing psychometric curves for congruent and incongruent audio-visual conditions (Figure 3.3A middle and right) showed that subjects preferentially weighted the auditory modality regardless of visual reliability, as demonstrated by shifts in the psychometric curves towards the auditory rate. However, as expected, this shift was more pronounced in the low reliability condition, indicating a stronger influence of the auditory modality when visual reliability was reduced (Table 3.1, Figure 3.3A, right).

Figure 3.3B displays the individual threshold performance (extracted from the psychometric curves) for each condition. On unisensory trials, thresholds were significantly lower (i.e. better performance) for high compared to low reliability stimuli across subjects ($p<0.05$, Table 3.1). Thresholds were comparable for the auditory and both congruent audio-visual conditions (all comparisons $p>0.05$, Table 3.1). Finally, thresholds on audio-visual trials were significantly lower compared to the visual conditions ($p<0.05$, Table 3.1). This demonstrates that performance was comparable for audio-visual and auditory trials, lowest for visual trials, and better for high vs. low reliable stimuli.

Figure 3.3C presents the predicted weights for each modality and reliability condition calculated using the optimal integration model. Predicted auditory weights significantly increased and visual weights decreased when visual reliability was reduced ($p<0.05$, Table 3.2). However, this was not consistently found in the observed weights (Figure 3.3D), and there was no significant difference between observed weights for reliabilities in each modality ($p>0.05$, Table 3.2). This suggests that individual subject weighting could not be adequately accounted for by the optimal integration model. This is corroborated by the magnitude and direction of the weight shift across subjects; only 11 subjects showed perceptual weight shifts in the predicted direction (i.e.

increased auditory weighting when visual reliability is reduced). Furthermore, a direct comparison between observed and predicted weights revealed only a weak correlation for low reliability weights ($p<0.05$, Table 3.2).

Overall, the lack of a significant difference between unisensory and multisensory thresholds and the weak correlation between observed and predicted weights suggests that observers did not systematically follow the behavioural pattern predicted by optimal models of multisensory integration. This heterogeneity in the change of perceptual weights with reliability presents a unique opportunity to investigate the neural correlates of perceptual weights independently of an effect of sensory reliability, as these two effects are dissociable across subjects.

Evolution of Perceptual Weights over Time

Figure 3.3E displays the results for classification of behavioural choice (first stream higher/lower than second) from accumulated rate. Sensory evidence (accumulated rate) was significantly predictive of behavioural choice across the trial (permutation test, $p<0.01$), and increased as the trial progressed. Figure 3.3F shows the time-resolved perceptual weights, and demonstrates that auditory and visual weights changed significantly with reliability early during the trial; two auditory clusters (cluster 1, 48 ms to 192 ms; cluster 2, 276 ms to 300 ms, Table 3.3) and one visual cluster (cluster 1, 48 ms to 216 ms, Table 3.3). Additionally the results confirmed that subjects preferentially weighted the auditory over the visual modality (cluster randomisation test, $p<0.05$, Figure 3.3F, Table 3.3). Finally, Figure 3.3G shows the correlation between these time-resolved perceptual weights and those derived from the psychometric curve analysis. Significant correlations emerged during three epochs that collectively covered most of the trial: 132-192 ms, 252-408 ms, and 420-600 ms (cluster randomisation tests, $p<0.05$, Table 3.3).

In summary, these behavioural results show: (a) that auditory weights increase as visual reliability decreases, (b) subjects show a bias towards the auditory modality over visual regardless of analysis method, and (c) that our two measures of perceptual weighting correlate highly with each other.

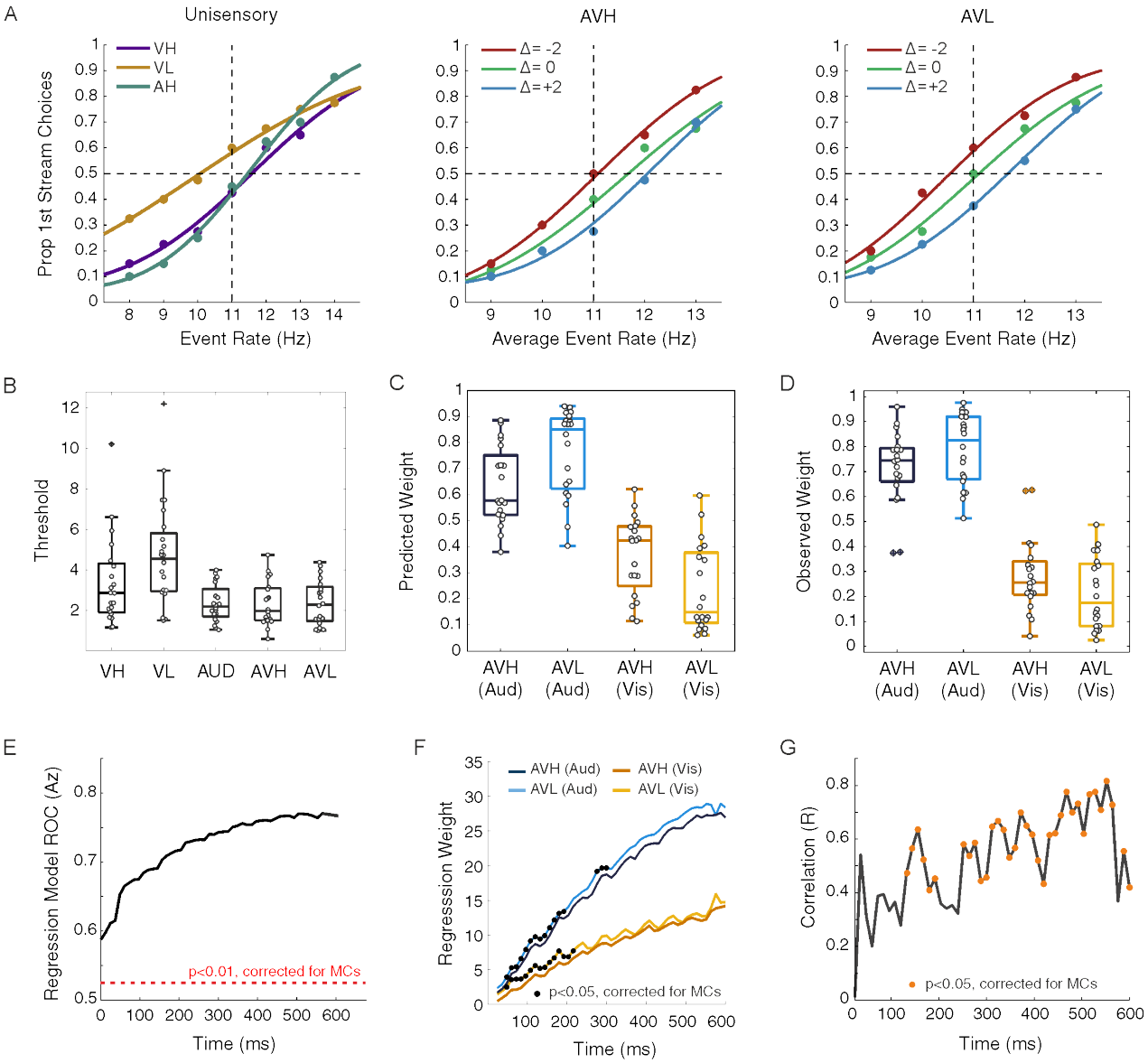


Figure 3.3 | Behavioural Results. (A) Group (n = 20) level psychometric curves displayed as the proportion of "first stream" decisions as a function of event rate for each condition. Note that for incongruent trials the x-axis indicates the average event rate (Δ = Visual Rate – Auditory rate). Vertical dashed lines represent the standard rate (11Hz) and horizontal dashed lines represent chance (50%) performance. (B) Individual subject threshold values (σ) for each separate condition, with group result shown on boxplot. (C) Group level predicted perceptual weights, with individual subject data shown in grey. AVH represents the audiovisual condition where both the auditory and visual cues were equally reliable. AVL represents the audiovisual condition where the auditory was highly reliable and the visual was less reliable. Within each AV(H/L) condition, auditory values are denoted with Aud, and visual values denoted with Vis. This gives four sets of weights, one for each modality within each reliability condition (AVH Aud, AVL Aud, AVH vis, AVL vis). (D) Observed Perceptual Weights, plotting same information as in (B). (E-F) Logistic regression was used to predict single trial choice (event rate >/<11Hz) based on the accumulated event rate at each time point in the trial. (E) Performance of the logistic model used to quantify how well our measure of sensory evidence (accumulated rate) predicted behavioural choice (first stream higher/lower), calculated using the area under the ROC and 10-fold cross validation (red dashed line p<0.01, randomisation test) (F) Auditory and visual perceptual weights derived from the regression model .Time points with significant reliability effects are denoted with black circles. (G) Correlation of perceptual weights derived from psychometric curves and from the logistic model. Time points with significant correlations are marked with orange circles.

Psychometric Fits			Comparison of Perceptual Thresholds			
	Median σ	Median μ		Z value	p value	Effect Size
AUD	2.19	11.45	VH vs AUD	-2.688	0.007	-0.425
VH	2.87	11.44	VL vs. AUD	-3.658	0.0003	-0.579
VL	4.55	10.53	VH vs. VL	-3.136	0.002	-0.496
AVH (Δ = 0)	1.98	11.72	AUD vs. AVH	-0.336	0.737	-0.053
AVH (Δ = +2)	1.80	12.25	AUD vs. AVL	-0.261	0.794	-0.041
AVH (Δ = -2)	1.89	11.05				
			AVH vs. AVL	-0.018	0.986	-0.003
AVL (Δ = 0)	2.29	11.14	AVH vs. VH	3.322	0.0009	0.525
AVL (Δ = +2)	1.82	11.94	AVH vs. VL	3.621	0.0003	0.573
AVL (Δ = -2)	2.10	10.58	AVL vs. VH	3.397	0.0007	0.535
			AVL vs. VL	3.919	0.00009	0.619

Table 3.1 | Analysis of Psychometric Data. Median threshold (σ) and PSE (μ) values from fits to psychometric data (left). Statistical tests (right) were based on two sided Wilcoxon Signed Rank tests of thresholds (σ).

High vs. Low Reliability Psychometric Weights				Observed vs. Predicted Psychometric Weights		
	Z	P	Effect		R	P
Visual (Pred)	-3.09	0.0019	-0.49	Visual (High)	0.18	0.45
Visual (Obs)	-1.64	0.1002	-0.26	Visual (Low)	0.46	0.04
Auditory (Pred)	-2.84	0.005	-0.45	Auditory (High)	0.21	0.381
Auditory (Obs)	-1.26	0.207	-0.20	Auditory (Low)	0.48	0.034

Table 3.2 | Analysis of Predicted and Observed Psychometric Weights. Comparison of psychometric weights based on two sided Wilcoxon signed rank tests (*left;* High vs. Low reliability comparisons) and Spearman rank correlations (*right*; R_s, Predicted vs. Observed comparisons).

Time-resolved Perceptual Weights					
	Cluster #	Time (ms)	p-value	t-value	Effect Size
AH vs AL	1	48 to 192	<0.0001	-13	0.5005
	2	276 to 300	0.010	-3	0.5604
VH vs VL	1	48 to 216	<0.0001	-15	0.5976
PMC vs PRW *	1	48 to 588	<0.0001	6	0.5201

Neural Weights: Modality Dominance					
	Cluster #	Time (ms)	p-value	t-value	Effect Size
AH vs VH	1	36 to 60	<0.001	3	0.5576
	2	108 to 120	<0.001	2	0.5055
	3	252 to 264	<0.001	-2	0.4566
AL vs VL	1	60 to 96	<0.001	11	0.5564
	2	120 to 336	<0.001	2	0.5014

Neural Weights: Sensory Reliability					
	Cluster #	Time (ms)	p-value	t-value	Effect Size
AH vs AL	1	156 to 204	<0.0001	-5	0.4940
	2	264 to 276	0.0070	-2	0.4731
VH vs VL	1	84 to 108	<0.0001	3	0.5139
	2	252 to 288	<0.0001	4	0.5406

Neural vs. Perceptual Weights					
	Cluster #	Time(ms)	p-value	t-value	Effect Size
NW vs PRW *	1	120 to 132	0.005	2	0.4674
	2	204 to 228	<0.0001	3	0.5170

Table 3.3 | Statistical Comparisons of Perceptual and Neural Weights. Tests were performed using cluster randomisation statistics (*see Statistics*). For each significant effect we list cluster p-value (where p-values below 10^{-3} are listed as <0.001), cluster t-values and effect size. T values marked with * reflect t-values derived from correlations. PMC = Psychometric Curve Weight. PRW = Perceptual Regression Weight. NW = Neural Weight. Condition abbreviations (AUD,VH,VL,AVH,AVL) see: Methods.

EEG Decoding Components and Neural Weights

Figure 3.4A displays the discriminant performance across subjects. Significant performance emerged early in the trial from 48 ms to 396 ms (permutation test, $p<0.01$) and was highest during three epochs: 96 ms to 120 ms, 168 ms to 204 ms, and 252 ms to 288 ms (Az performance >0.58). This demonstrates that there was information related to stimulus rate (high/low) evident in EEG signal early in the trial. The corresponding scalp topography (Figure 3.4A, insert) indicated that the strongest information about audio-visual rate emerged over posterior and central electrodes, consistent with an origin in early sensory (temporal and occipital) cortices.

Figure 3.4B displays the neural weights derived from the regression analysis. Neural weights exhibited a similar bias towards the auditory modality as perceptual weights; auditory weights dominated in both reliability conditions (Figure 3.4B, left and right). When both auditory and visual stimuli were highly reliable (audio-visual high reliability condition), auditory weights were higher than visual at three separate clusters (cluster 1: 36 ms to 60 ms, cluster 2: 108 ms to 120 ms, cluster 3: 252 ms to 264 ms; cluster randomisation tests, $p<0.05$, Table 3.3). When the auditory stimuli were highly reliable and visual stimuli were not (audio-visual low reliability condition), auditory weights were higher than visual at two clusters (cluster 1: 60 ms to 96 ms, cluster 2: 120 ms to 336 ms; cluster randomisation tests, $p<0.05$, Table 3.3). Consistent with perceptual weights, auditory weights were significantly higher when the visual reliability was reduced (compared to auditory weights when visual reliability was high) and these differences emerged during two epochs (156 ms to 204 ms, and 264 ms to 276 ms; cluster randomisation test, $p<0.05$; Figure 3.4C left; Table 3.3). Visual weights were significantly lower when the visual reliability was reduced (compared to visual weights when visual reliability was high) at two epochs (84 ms to 108 ms, and 252 ms to 288 ms; cluster randomisation test, $p<0.05$, Figure 3.4C right, Table 3.3). Finally, there was a significant correlation between the reliability effect on the time-resolved perceptual and the neural weights at two epochs (120 ms to 132 ms, and 204 ms to 228 ms (cluster randomisation test, $p<0.05$, Figure 3.4D, Table 3.3).

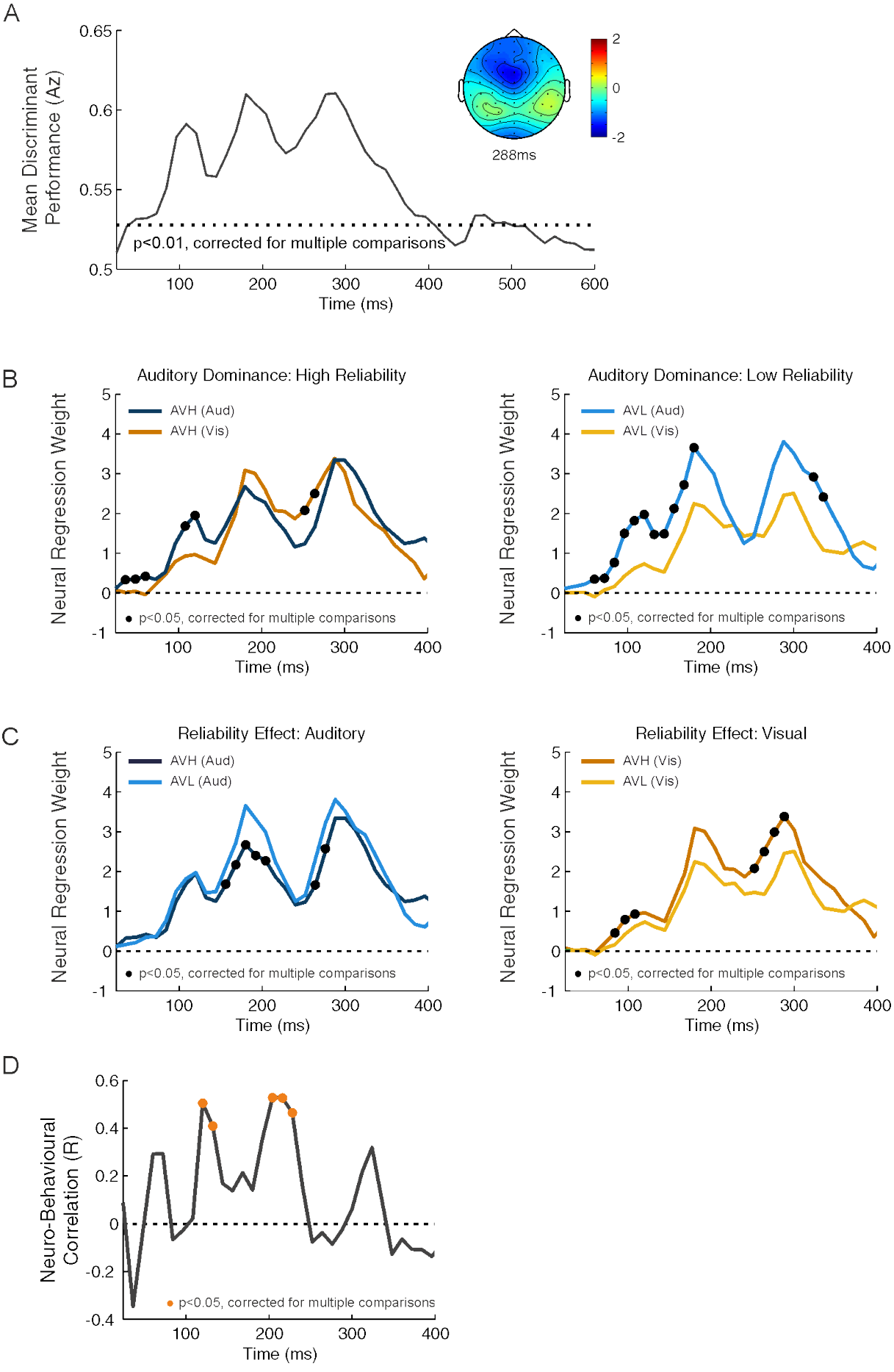

A
Mean Discriminant Performance (Az)
0.65
0.6
0.55
0.5
2
0
-2
288ms
p<0.01, corrected for multiple comparisons
100
200
300
400
500
600
Time (ms)
B
Auditory Dominance: High Reliability
Neural Regression Weight
AVH (Aud)
AVH (Vis)
p<0.05, corrected for multiple comparisons
Time (ms)
Auditory Dominance: Low Reliability
AVL (Aud)
AVL (Vis)
C
Reliability Effect: Auditory
Reliability Effect: Visual
D
Neuro-Behavioural Correlation (R)
0.6
0.4
0.2
-0.2
-0.4

Figure 3.4 | Decoding, Neural Weights and Neuro-Behavioural Correlation. (A) Group averaged performance of a linear classifier applied to all audio-visual congruent trials. Classifier trained to discriminate between event rate >/<11Hz (i.e. stimulus rate), quantified using the area under the ROC curve. Scalp topography (insert) represents the forward model (correlation between discriminant output Y and underlying EEG activity) for the peak decoding time point (288 ms). Colorbar represents correlation value. (B) Neural weights for each modality for high reliability (left) and low reliability (right) trials, shown for each modality separately (auditory = blue, visual = yellow). (C) Neural weights in each reliability condition for auditory (left) and visual (right) trials, shown for each reliability level separately (high reliability = darker colours, low reliability = lighter colours in each graph). In (B,C), time points with significant effects are indicated by black circles. (D) Neuro-behavioural correlation between perceptual and neural weights obtained from the regression models. Time points with significant correlations indicated by orange circles.

Summarising these results in order of time (rather than by statistical contrast) reveals an evolving pattern of results as the trial progresses. From stimulus onset, first there was a change in visual weights (starting 84 ms) and a significant relationship between perceptual and neural weights (starting 120 ms). This was followed by a change in auditory weights (starting 156 ms) and another epoch where there was a significant relationship between perceptual and neural weights (starting 204 ms). Finally, there was a change in both visual (starting 252 ms) and auditory weights (starting 264 ms) later in the trial. This indicates a dynamic process whereby sensory information is weighted in the brain.

To disentangle whether these six epochs represented distinct neural processes or whether they related to the same underlying neural generators, we compared the scalp projections and neural weights between the six epochs (Supplementary Figure S3.2). This revealed that temporally adjacent topographies (84 to 108 ms and 120 to 132 ms; 156 to 204 ms and 204 to 228 ms; and 252 to 288 ms and 264 to 276 ms) were highly correlated (within Epochs: $R_S > 0.6$, $p < 0.005$). The reliability difference in neural weights (Equation 3) at temporally adjacent epochs were also highly correlated ($R_S > 0.6$, $p < 0.001$) and showed similar patterns of neural weights. For all other comparisons, the correlation values were all less than 0.6, and above the threshold value of 0.005. Consequently, we concatenated the six epochs into three separate epochs of interest based on their high correlations (Epoch 1: 84 ms to 132 ms; Epoch 2: 156 ms to 228 ms; and Epoch 3: 252 ms to 276 ms).

Localization of Neural Sources

Figure 3.5 shows the neural weights, forward model scalp topographies, and source localisation maps for each of the three clusters defined above. Table 3.4 reports co-ordinates and statistical values for significant voxels. The first epoch (84 ms to 132 ms) was characterized by a scalp projection consistent with a potential origin in sensory cortices. Source localisation revealed that discriminant activity originated from occipital and temporal regions. The second epoch (156 ms to 228 ms) had a scalp projection that revealed strong contributions from fronto-central and occipital electrodes, and source activity was broadly localized to temporal, occipital and parietal regions. The third epoch revealed contributions from central electrodes with sources in occipital and parietal regions.

Epoch 1: 84 ms to 132 ms					
	Co-ord (spm)	z-transformed R	p-value	t-value	Effect Size
Occipital Mid L	(-45,-70,0)	0.145	0.004	-6.949	0.486
Temporal mid L	(-60,-38,-10)	0.197	0.004	-7.687	0.526
Epoch 2: 156 ms to 228 ms					
	Co-ord (spm)	z-transformed R	p-value	t-value	Effect Size
Occipital Mid R	(37,-72,27)	0.266	<0.001	-8.336	0.541
Temporal Inf. R.	(58,-15,-25)	0.267	<0.001	-6.768	0.476
Parietal Inf. R.	(47,-57,41)	0.268	<0.001	10.275	0.439
Epoch 3: 252 ms to 276 ms					
	Co-ord (spm)	z-transformed R	p-value	t-value	Effect Size
Occipital Mid. L.	(-35,-86,18)	0.166	0.003	7.575	0.416
Parietal L.	(-36,-32,46)	0.159	<0.001	12.223	0.499

Table 3.4 | Source Localisation of Discriminant Activity. Significance values obtained from correlating source signals with discriminant output (Y), corrected for multiple comparisons using cluster permutation testing (see Statistics). For each significant source the table provides co-ordinates of peak voxel (co-ord spm), z-transformed R value (as plotted in Fig.3.5, bottom row), and the p-value, cluster t-value, and effect size.

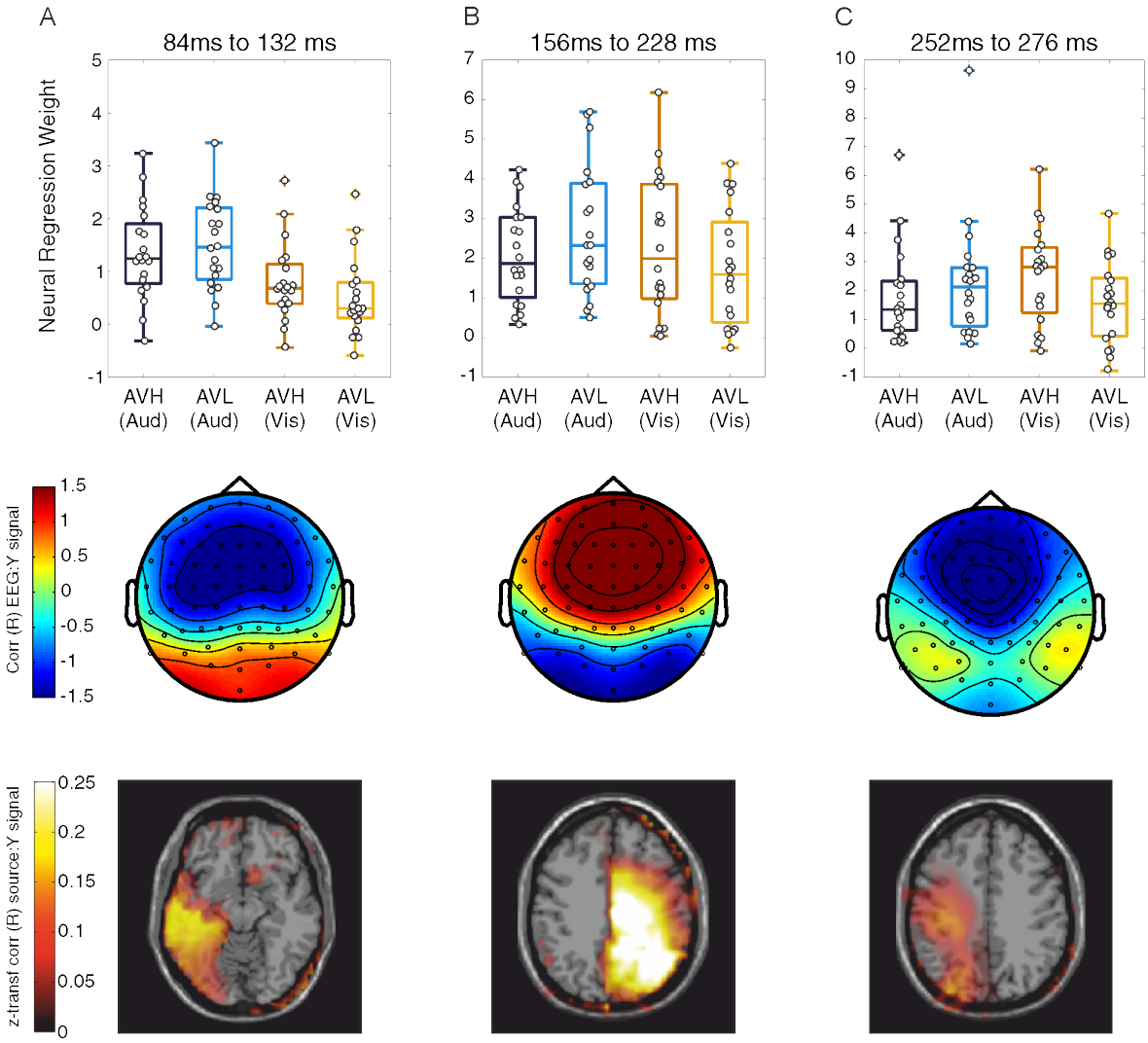


Figure 3.5 | Neural Weights, Topographies and Source Localisation results for three EEG components of interest. Each cluster was defined based on the statistical contrast between sensory reliabilities (Fig. 3.4C), or a significant neuro-behavioural (N2B) correlation (Fig. 3.4D). In each panel, boxplots represent neural weights averaged over each epoch, with individual subject data in grey. Topographies represent the group averaged forward models averaged over the epoch (values represent correlation between discriminating output (Y) and underlying EEG activity). Source maps represent the z transformed correlation values of single voxel activity with the discriminant output for each epoch (left hemisphere on left-hand side. Source localisation z-coordinates for slice: (A) -7, (B) 37, and (C) 51).

Discussion

This study examined the temporal dynamics underlying audio-visual cue weighting in the brain. Our behavioural results revealed that subjects weighted sensory cues in proportion to their relative reliabilities: auditory perceptual weights increased as visual reliability decreased. These modulations in behavioural reweighting occurred early during the trial (starting 48 ms). We also found a consistent bias towards the auditory modality regardless of visual reliability, which caused a mismatch in the perceptual weights predicted by the optimal integration model and those observed in subject behaviour. The EEG data revealed that information related to stimulus type could be reliably extracted from neural signals during stimulus presentation. Additionally, we found three early epochs (starting 84 ms) during which EEG correlates were modulated by sensory reliability. Similarly early neural correlates of perceptual weighting were also evident early in the trial, emerging 120 ms after stimulus onset. However, these two processes had distinct topographical patterns, and source localisation revealed these neural correlates of sensory reliability and perceptual weighting localised to early sensory (occipital, temporal) and parietal regions respectively. Taken together, these results shed new light on the temporal dynamics of underlying audio-visual cue weighting, and suggest that reliability encoding and perceptual reweighting are computationally distinct processes in the brain.

Perceptual Weights: Evolution over Time

We performed two analyses to generate perceptual weights. The first followed the conventional approach using the optimal integration framework to calculate weights based on psychometric curves fit to subjects' responses. The second used regression modelling to generate a set of perceptual weights based on the time-varying accumulated sensory evidence. This dual approach allowed different views of the reliability effects on behavioural weights, and the novel regression approach gave us an opportunity to apply comparable analyses to behavioural and neural data (as in Fetsch et al., 2012; Gu et al., 2008).

However, these two approaches revealed divergent results regarding the effect of sensory reliability on perceptual weights. The analysis of psychometric curves

revealed no significant group effect of reliability, while the time-resolved behavioural weights were significantly modulated by reliability in the direction as expected by previous literature (that is, auditory weights increased as visual weights decreased). One potential explanation for this apparent discrepancy is methodological. In the optimal integration analysis, psychometric performance is measured only at the end of the trial and is based on trial labels. In contrast, the regression analysis quantifies choice at each time point based on the time-aggregated sensory evidence. Hence, one possibility is that early sensory information may contribute to perceptual weighting in a more specific manner than the average sensory information available throughout the trial, and so this effect of sensory reliability emerges only in the time-resolved weights. Despite this divergence, our finding of a significant correlation between the reliability effects for each set of perceptual weights suggests there is overall similarity between the two sets, and thus we feel confident that conceptually they measure the same process. This interpretation could be checked by reverse-correlating behavioural choice with the local rate (i.e. the momentary rate, not the accumulated rate) across time, especially for those trials where the comparison rate across streams is equal (11Hz – 11 Hz trials).

Perceptual Weights: Auditory Bias

Our group data demonstrate that subjects failed to show the expected multisensory benefit, as shown by comparable performance on audio-visual and auditory trials. We also observed a general bias towards the auditory stimulus regardless of visual reliability in the majority of subjects. This bias emerged despite efforts to equalise visual and auditory thresholds using unisensory calibration blocks.

With regards to this experiment, we hypothesise that the observed auditory bias arises from the preference for auditory over visual information for temporal judgements (Glenberg & Fernandez, 1988; Glenberg, Mann, Altman, Forman, & Procise, 1989; Recanzone, 2003; Repp & Penel, 2002). As discussed in the general introduction chapter (Chapter 1), the auditory modality is more suited to temporal tasks while the visual modality is more suited to localisation tasks. Given that our task required rate discrimination (a temporal task), it could be that overall subjects found it easier to judge the stimulus rate on auditory trials

compared to visual trials regardless of reliability. Alternatively, it could be that as the auditory stimulus was presented in silence while the visual stimulus was embedded in noise, there may have been subtle differences in intramodal attention which caused a bias towards the auditory modality (Alho, Woods, Algazi, & Näätänen, 1992; Lu & Dosher, 1998). A final possibility for the observed auditory bias is that subjects did not consider auditory and visual stimuli to originate from the same underlying sensory cause, in particular on trials where event rates differed. Such stimulus-dependent changes in the inference about the causal structure of the environment have recently been included in models of sensory integration (Shams et al., 2005; Roach et al., 2006; Knill, 2007; Körding et al., 2007; Beierholm et al., 2009; for review, see Shams & Beierholm, 2010 and Kayser & Shams, 2015), and recently neuroimaging studies have started to study the neural mechanisms underlying this flexibility in sensory integration (Rohe & Noppeney, 2015a, 2016). However, given that the present experiment included only one level of audio-visual discrepancy it is not possible to ascertain whether causal inference processes contribute to the apparent mismatch between the observed psychometric performance and predictions based on optimal integration strategies.

Interestingly, similar modality biases have been reported in the cue weighting literature before (Fetsch et al., 2012, 2009; Knill & Saunders, 2003; Oruç, Maloney, & Landy, 2003; Rosas et al., 2005; Rosas, Wichmann, & Wagemans, 2007; Sheppard et al., 2013), and attempts to improve the predictions of optimal integration models have been made. For example, Battaglia et al., (2003) found that both the reliability of a visual stimulus and a bias for visual information over auditory affected the perceptual weights in a spatial localisation task. They showed that a “hybrid model” which included a prior to make greater use of the visual information provided a better fit to the data than the standard optimal integration model. Butler et al., (2010) found similar results for visual-vestibular stimuli. Taken together, these studies demonstrate that including a prior to account for modality bias can improve the predictions of optimal integration models. Adapting the integration model to include such a prior was beyond the scope of this project, but provides an interesting starting point for future work.

Decoding Correlates of Stimulus Rate from EEG Activity

Our results showed we were successfully able to extract information relating to stimulus rate from the EEG signal early during stimulus presentation. This adds to the growing literature which has used multivariate single trial decoding of EEG data to reveal the dynamic correlates of various types of sensory stimuli (Kayser et al., 2016; Lou, Li, Philiastides, & Sajda, 2014; Mostert, Kok, & de Lange, 2015; Philiastides et al., 2014; Philiastides & Sajda, 2006; Ratcliff et al., 2009; Wyart, de Gardelle, Scholl, & Summerfield, 2012). In addition, we exploited this approach to quantify multisensory interactions directly within those EEG components carrying the relevant sensory information, rather than relying on generic stimulus-related evoked responses.

However, our results demonstrated that the classifier was not able to successfully decode stimulus rate beyond 400 ms after stimulus onset. There are two possible explanations for this: either sensory rate is only linearly reflected in EEG activity early during the trial, or neural activity later in the trial reflects a mix of sensory encoding and decision-making processes which may make it difficult to extract purely sensory representations (Raposo, Kaufman, & Churchland, 2014). While this precludes us from making statements about the patterns of sensory weighting that may occur later in the trial, our results directly reveal neural correlates of changes in sensory reliability and of perceptual weights that are evident early during the integration process.

Early Neural Correlates of Sensory Reliability and Perceptual Weights

We found neural correlates of both sensory reliability and perceptual weights at multiple times early points. We thereby step beyond previous neurophysiological (Fetsch et al., 2012; Gu et al., 2008; Morgan et al., 2008) and neuroimaging studies (Beauchamp et al., 2010; Helbig et al., 2012; Rohe & Noppeney, 2015a, 2016) to reveal the temporal evolution of the sensory weighting process in functionally specific brain activity. In addition, by dissociating the influence of sensory reliability on the representation of sensory information from perceptual weighting in EEG responses rather than demonstrating a simple modulation of evoked response amplitudes, we demonstrate that these early effects reflect sensory and computationally specific processes. This complements recent work

in the fMRI literature by Rohe & Noppeney, (2016) which found that sensory reliability was encoded in early sensory areas while higher-order parietal areas were responsible for weighting based on stimulus reliability. Thus our results provide support for computationally distinct processes occurring across the cortex to implement cue weighting in the brain, but add the important temporal dimension that up until now has been missing.

Early Effects of Sensory Reliability

We found that neural correlates of cumulative sensory evidence scaled with reliability. At the earliest time (84 ms) these effects were associated with changes in visual weights, while at a slightly later window (starting at 156 ms) these effects were associated with changes in auditory weights. Even later in the trial (starting at 252 ms), changes in both auditory and visual weights were evident.

First, the early onset of these changes in audio-visual sensory weights supports the notion of low-level and short-latency multisensory interactions (Cappe et al., 2010; Foxe et al., 2000; Giard & Peronnet, 1999; Molholm et al., 2002; Murray et al., 2005; Sperdin et al., 2009). Second, our finding that visual and auditory weights scaled with reliability at different latencies during the trial is noteworthy. While visual weights were affected early (<100 ms), auditory weights increased with decreasing reliability of the visual stimulus later (around 150 ms). This temporal dissociation between visual and auditory weights shifts could reflect the adaptive nature of multisensory integration; perhaps visual encoding is adjusted at short latencies and in a bottom-up (i.e. sensory driven manner) to cope with trial-by-trial changes in visual sensory reliability. In contrast, auditory encoding may be adjusted only later (possibly as result of top-down processes) in order to meet the increased demands for representing the unreliable sensory environment. Hence, it makes sense for the brain to first adapt to the visual modality on a trial-by-trial basis, before subsequently adjusting processing across the auditory modality. This speculation is in part supported by the fact that auditory reliability was fixed throughout the experiment, while visual reliability varied unpredictably.

Early Neural Correlates of Perceptual Weighting

As not every subject attributed perceptual weights in a statistically optimal manner, we were able to dissociate neural correlates related to perceptual weighting from correlates related to sensory reliability. First, this demonstrates that the scaling of sensory representations based on stimulus reliability and the process of perceptual weighting are computationally distinct and reflect different aspects of the perceptual process. Second, we found that neural correlates of the perceptual weighting process emerged early in the trial, at 120 ms and 204 ms after stimulus onset. These effects in neural signals – while somewhat later than the onset of modulations in the behaviour – are still earlier than expected, and a long time before the perceptual choice at the end of the trial. Furthermore, the early onset of these effects suggests these neural correlates are not related to the perceptual decision.

Supporting this, past work specifically aiming to dissociate neural correlates related to sensory processing from those related to perceptual decisions has shown that these two processes have different temporal profiles. For example, Mostert et al., (2015) showed that sensory information was encoded from 130 ms, while decision related processes emerged later from 250 ms to 600 ms. Wyart et al., (2012) found similar results, showing sensory correlates peaked around 120 ms while perceptual decision correlates emerged around 300 ms. In our case, we speculate that the early onset of our effects (i.e. all our effects emerged before 300ms) suggest that perceptual weights are adjusted on each trial individually at early sensory stages to allow the information to be used in the later perceptual decision. What remains unclear is whether these perceptual weights are adjusted on each trial individually and in response to the experienced sensory reliabilities, or whether they are at least in part already established based on task-context in a predictive manner even before stimulus onset. Future work is required to elucidate the precise neural correlates of these perceptual weights, and examine how different brain regions contribute to establishing the perceptual integration process.

Localising these Effects in Space and Time

Finally, the temporal organization and localisation of the reliability and perceptual weighting effects in three clusters showed distinct patterns of topographies and neural sources. At the earliest window (84 ms to 132 ms), effects of sensory reliability and perceptual weights were associated with occipital scalp topographies and source activity emerging from early visual and temporal areas. At the slightly later time point (starting at 156 ms to 228 ms), effects were associated with activity over central electrodes (consistent with activations including prominent contributions from auditory cortex) and source activity from temporal and parietal regions. Finally, at the latest window (252 ms to 288 ms), effects were associated with activity over posterior and central regions, and again source activity indicating a potential origin in occipital and temporal regions. While the source localization of the relevant EEG components was quite distributed, our results fit with the notion that earliest effects arise from occipital sensory regions and are followed by activity in the temporal and parietal lobe.

This evolving pattern of activation demonstrates early sensory and parietal regions encode sensory cues and represent the integrated evidence weighted by the relative reliability. This complements existing findings from fMRI work which have shown multisensory interactions occurring along primary sensory and parietal areas in response to changing reliability. For example, Helbig et al., (2012) found that BOLD responses in both primary somatosensory and the superior parietal lobe increased when visual reliability decreased during a visual-tactile task. Beauchamp et al., (2010) demonstrated that the strength of functional connections increased between somatosensory and intraparietal sulcus (IPS) for reliable somatosensory stimuli, but increased between visual and IPS for more reliable visual stimuli. Finally, Rohe & Noppeney (2015a, 2016) showed that primary sensory areas encoded the spatial location of cues during an audio-visual task, while early parietal areas (IPS1-2) represented the reliability weighted signals. Taken together with the prior literature, our results support the idea that sensory reweighting is an evolving and hierarchical process, with multisensory interactions emerging along the sensory pathway in primary sensory and parietal areas. Yet our results add a temporal dimension to these processes and demonstrate that effects related to external sensory

reliability and perceptual weighting emerge at slightly different times and from distinct brain regions.

Chapter 3: Summary

This chapter has presented results showing that neural correlates of sensory reliability and perceptual weights at multiple times, early during a trial. This was achieved using single-trial, time-resolved analysis of neural and behavioural data, thereby stepping beyond previous neurophysiological (Fetsch et al., 2012; Gu et al., 2008; Morgan et al., 2008) and neuroimaging studies (Beauchamp et al., 2010; Helbig et al., 2012; Rohe & Noppeney, 2015a, 2016). In addition, by dissociating the influence of sensory reliability on the representation of sensory information from perceptual weighting in EEG responses rather than demonstrating a simple modulation of evoked response amplitudes, we show that these early effects reflect sensory and computationally specific processes. Overall, these results provide the first insights into the temporal evolution of cue weighting in the brain, thus adding an important contribution to the field. Additionally, the results provide the first evidence from this thesis supporting the modern view that multisensory interactions can occur early and across multiple regions of the perceptual pathway, and showing that neural and behavioural signals can be linked in a meaningful way.

Supplementary Materials: Chapter 3

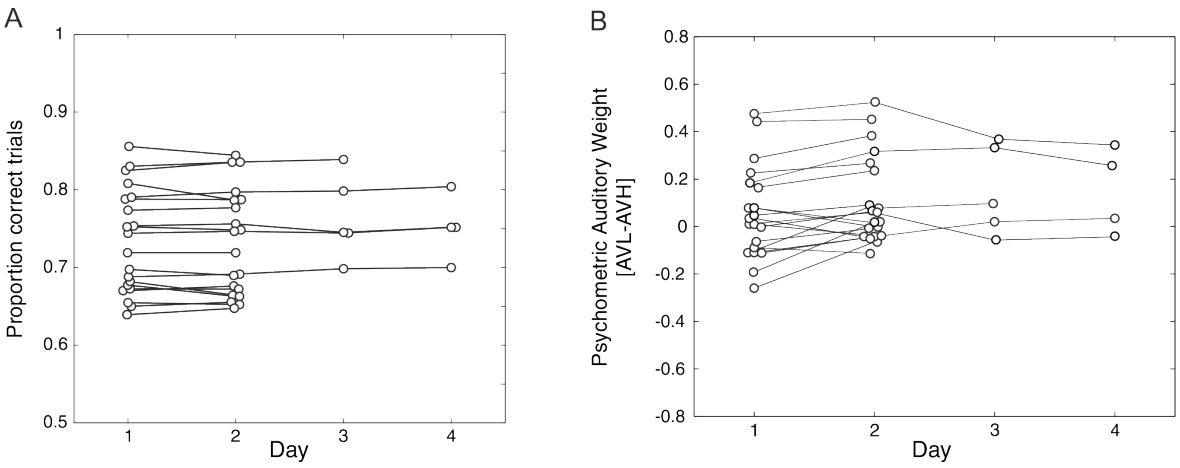


Figure S3.1 | Performance and Perceptual Weighting across Days. A) Single-subject performance (overall performance score) on each experimental session. B) Single-subject auditory weight difference (auditory low – auditory high) on each experimental session.

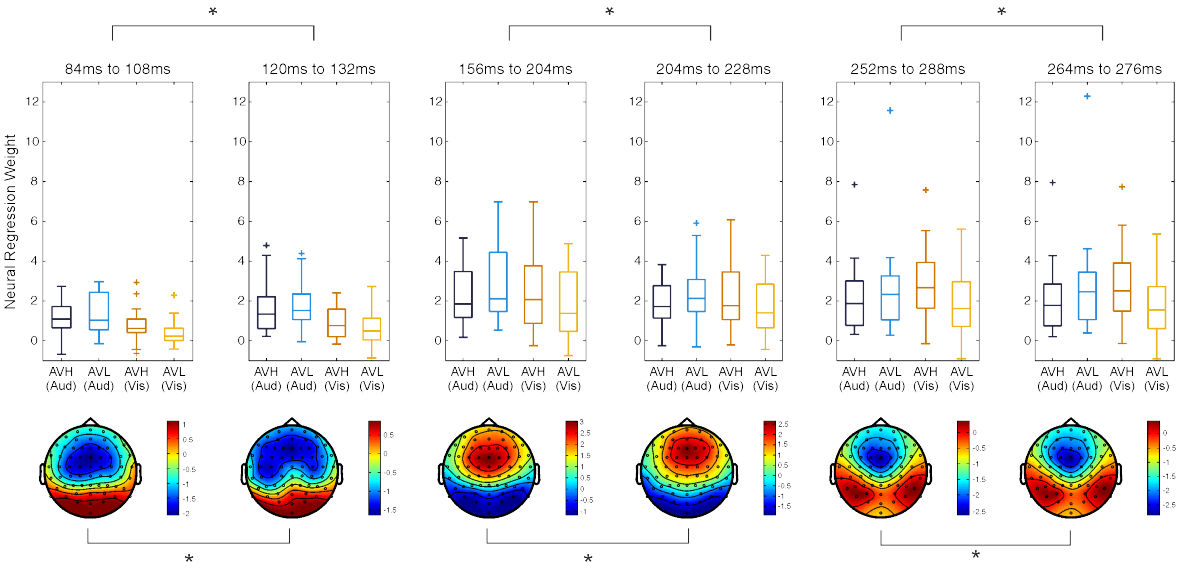


Figure S3.2 | Neural Weights and Scalp Topographies. Weights and topographies underlying the six time epochs that showed a significant effect of reliability. Topographies marked with * represent R >0.6 and p<0.005 for correlations between scalp topographies. For each epoch the difference in neural weights (AVH_{AUD} - AVH_{VIS}) - (AVL_{AUD} - AVL_{VIS}) was calculated and correlated, * represent R>0.6 and p<0.001.

	Day 1			Day 2			Day 3			Day 4		
	AUD	VH	VL	AUD	VH	VL	AUD	VH	VL	AUD	VH	VL
S01	83.1	26.2	14.5	80.2	34.3	12.7						
S02	92.0	32.6	10.2	97.1	51.1	27.4	98.3	33.4	10.9			
S03	78.3	74.2	20.8	96.6	70.1	22.3	98.3	29.5	12.4	90.2	76.0	23.8
S04	93.3	25.6	12.7	75.0	11.8	9.4	91.6	59.1	19.6	96.6	80.0	29.5
S05	86.7	25.3	10.9	80.0	89.0	22.0	95.0	70.0	19.0	90.0	56.2	13.3
S06	88.3	33.1	13.3	93.3	37.0	20.5	85.0	29.2	16.0	95.0	82.0	11.5
S07	98.3	64.3	18.4	94.2	84.0	19.6	100	40.0	11.2			
S08	76.7	15.3	11.1	73.3	25.6	10.0						
S09	86.6	33.1	10.6	88.3	70.0	40.0						
S10	76.6	71.6	11.8	66.6	12.7	8.4						
S11	71.6	13.9	7.0	91.6	87.0	25.0						
S12	85.0	48.4	23.2	91.6	65.8	18.7						
S13	85.0	78.2	13.3	73.3	22.6	7.0						
S14	98.3	25.0	11.5	100	40.6	11.8						
S15	73.3	47.2	13.6	88.3	71.0	16.0						
S16	85.0	71.0	29.8	81.6	25.0	10.3						
S17	90.0	70.0	19.6	88.3	10.3	4.0						
S18	98.3	73.7	16.3	97.6	43.9	16.9						
S19	83.3	25.6	14.0	90.0	81.0	17.8						
S20	84.3	12.6	6.0	85.0	41.2	19.6						

Table S3.1 | Calibration Block Thresholds. AUD column contains overall auditory performance score (% correct). VH and VL columns contain threshold values (contrast value) for signal to noise ratio (SNR) for high and low reliability visual stimuli respectively.

Chapter 4 : Theta and Alpha Power are modulated by the Reliability of Sensory Information Early during Audio-visual Integration.

Introduction

Brain signals show rhythmic fluctuations. These fluctuations, known as "neural oscillations", have been shown to underlie many perceptual and cognitive functions (Başar, Başar-Eroglu, Karakaş, & Schürmann, 2000; Buzsáki & Draguhn, 2004; Siegel, Donner, & Engel, 2012), including multisensory perception. For example, low frequency activity delta (~1-4 Hz), theta (~4-8 Hz), alpha (~8-12 Hz) and beta (~13-30 Hz) activity have been found during human audio-visual paradigms (Luo, Liu, & Poeppel, 2010; Senkowski, Molholm, Gomez-Ramirez, & Foxe, 2006; Thorne, De Vos, Viola, & Debener, 2011; Sakowitz, Schürmann, & Başar, 2000), as well as in direct neuronal recordings in animals studies (Kayser et al., 2008). High frequency (>30 Hz) gamma activity has been found in audio-visual speech paradigms (Doesburg, Emberson, Rahi, Cameron, & Ward, 2008; Kaiser, Hertrich, Ackermann, & Lutzenberger, 2006; Kaiser, Hertrich, Ackermann, Mathiak, & Lutzenberger, 2004), audio-visual semantic matching tasks (Schneider et al., 2008), and in response to synchronous audio-visual stimuli (Bhattacharya, Shams, & Shimojo, 2002; Maier, Chandrasekaran, & Ghazanfar, 2008; Widmann, Gruber, Kujala, Tervaniemi, & Schröger, 2007; Yuval-Greenberg & Deouell, 2007). Some studies have even demonstrated activity in both high and low frequency bands co-occurring during audio-visual listening tasks (Sakowitz, Quiroga, Schurmann, & Basar, 2005) and audio-visual speech paradigms (Arnal et al., 2011). Oscillations are in fact such a pervasive finding in the multisensory literature that they have been proposed as the potential mechanism used by the brain to integrate and combine information across different modalities (Buzsáki & Draguhn, 2004; Engel et al., 2012; Fries, 2005; Senkowski, Schneider, Foxe, & Engel, 2008).

While the evidence strongly suggests that oscillations play an important role in multisensory perception, there is still no consensus on what role that is, and there are many multisensory process yet to be explored. For example, the role neural oscillations play in the process of "optimal integration" has not yet been

investigated. As introduced in Chapter 3, the optimal integration framework states that observers can arrive at the most precise final perceptual estimate by weighting sensory cues in proportion to their relative reliabilities. Using this strategy, more reliable sensory cues have a greater influence and lead to the most precise final perceptual estimate. In support of this, many behavioural (Battaglia et al., 2003; Butler et al., 2010; Fetsch et al., 2009; Helbig & Ernst, 2007; Raposo et al., 2012; Sheppard et al., 2013), neuroimaging (Beauchamp et al., 2010; Helbig et al., 2012; Rohe & Noppeney, 2015b, 2016), and electrophysiological studies (Fetsch et al., 2012; Gu et al., 2008; Morgan et al., 2008) have found behavioural and neural evidence of cue weighting, thus suggesting it is an important mechanism of multisensory integration. Given that both oscillations and optimal integration are cited as important mechanisms for sensory integration, an investigation of the role (if any) oscillatory activity plays in cue weighting could be fundamental to gaining a complete understanding of multisensory integration in the brain.

To this end, in this study we investigated how oscillatory power was modulated by visual reliability during an audio-visual task. To do so, the EEG and behavioural datasets collected for Chapter 3 were re-analysed using a time-frequency approach and combined with behavioural regression analysis to link changes in power to changes in behaviour. Based on the findings from Chapter 3 (i.e. early onset effects) we hypothesised that any modulations in oscillatory power would occur early during the perceptual process. However, due to the literature showing oscillations in all bands emerging during audio-visual paradigms, we had no hypothesis about which frequency band would show modulations.

Methods

Stimuli, Participants, Task and EEG recording

The stimuli, task, EEG recording and preprocessing settings were as described in Chapter 3. To avoid excessive repetition, this section will only provide details of the new analysis. For a detailed overview please refer to Chapter 3.

Behavioural Analysis

Time-varying perceptual weights were calculated using the method described in Chapter 3. Please refer to “Chapter 3: Time Dependent Perceptual Weights”, and see “Chapter 3, Figure 3.2” for a visual overview of the analysis pipeline.

Time-Frequency Analysis

For each subject, time-frequency representations (TFR) were obtained using complex wavelet analysis alongside Hanning tapers (as implemented in the FieldTrip Toolbox, Oostenveld et al., 2011). TFR were calculated from -500 ms pre stimulus onset to 1200 ms post stimulus with a 50 ms sliding time window. Wavelet transforms (“mtmconvol” method) were calculated for all sensors in 1 Hz steps for frequencies between 1Hz and 10Hz, 2 Hz steps from 12 Hz to 24 Hz, and 4 Hz steps from 26 Hz to 80 Hz. To account for temporal smoothing, the width of individual wavelets was scaled (5 cycles per time window) for all frequency bands. For higher frequencies (>26 Hz), frequency smoothing was also applied (0.4 Hz smoothing) to induce stronger smoothing at higher frequencies. This resulted in two separate TFRs, one for low frequency and one for high. These two TFR were then appended to create one power spectrum (for each subject), which was then normalised relative to a baseline period of -500 ms to -100 ms (baseline normalisation as implemented in Fieldtrip, with setting “relative” dividing the data by the mean of data in baseline interval). All resulting power spectra represent baseline-normalised relative power.

Our main goal was to examine whether changes in visual reliability would modulate auditory power during audio-visual conditions. To do this, the individual subject power spectra were split into five different conditions based on reliability and modality: auditory (AUD), visual high (VH), visual low (VL), audio-visual high (AVH, where both the auditory and visual were highly reliable) and audio-visual low (AVL, where the auditory was of high reliability and the visual was of low reliability). Then the power related to the visual information was subtracted from each audio-visual condition separately using [AVH - VH] and [AVL - VL]). In theory, this subtracts all visual activity during audio-visual trials, and the remaining activity represents the auditory information elicited by audio-visual stimuli during each level of reliability.

To examine how auditory power was modulated by visual reliability during audio-visual trials, the two power spectra were tested against each other across subjects ([AVL - VL] versus [AVH - VH]). Statistical analysis was carried out using cluster based, non-parametric permutation analysis as implemented in FieldTrip (Maris & Oostenveld, 2007; Oostenveld et al., 2011). As our hypothesis was that modulations in relative oscillatory power would occur early during audio-visual integration, we set our time window of interest from 0 ms (stimulus onset) to 300 ms (the epoch during which effects of cue weighting emerged in Chapter 3). We had no a-priori hypothesis as to which frequency bands (if any) would show effects; consequently, we tested for effects in four separate frequency bands, defined here as: theta (4–8 Hz), alpha (8–12 Hz), beta (14–32 Hz), and gamma (30–50 Hz). For each comparison, frequencies within each band were averaged over and all time points and all electrodes were tested. Neighbouring electrodes were defined using a custom template based on Biosemi 64-channel layouts. For all comparisons, a dependent samples t-test was used, with significance level set to alpha = 0.025 (to control for the two sided test). Correction for multiple comparisons was performed using 1000 permutations, minimum cluster channel of 2, and the Monte Carlo method as implemented in Fieldtrip.

Finally, we examined the relationship between behavioural and oscillatory modulations. To do so, the power within each cluster showing significant effects was extracted and averaged over the significant frequency bands, time points, and electrodes for each reliability condition separately ([AVH - VH] and [AVL - VL]). This resulted in two power values for each subject (one in each reliability condition). Perceptual weights were averaged over the time windows of the significant clusters in the same manner as the oscillatory power for each reliability condition separately, resulting in two perceptual weight values for each subject.

Finally, the difference between high and low reliability trials ([AVL - VL] - [AVH - VH]) was calculated for power and perceptual weights for each individual subject separately, and the correlation between the two (power and perceptual weights) across subjects calculated using spearman-rank correlation analysis.

Results

Figure 4.1 shows the group-averaged power difference in auditory activity during high reliability audio-visual trials, relative to low reliability audio-visual trials (as calculated using [AVL - VL] – [AVH - VH]), for each frequency band separately (see Supplementary Figure S4.1 for individual power spectrums). Visual inspection of these power spectra indicated potential differences in theta, alpha, and beta power between reliability conditions. In the theta band, there appeared to be a fluctuation from lower power early in the trial to higher power later in the trial during audio-visual conditions when visual reliability was low, relative to trials where visual reliability was high. This was evidenced by negative values during the first half of the trial and positive values in the later half (as the difference is calculated as low reliability minus high reliability). In the alpha band it appeared there was higher power over frontal electrodes early in the trial for audio-visual trials where visual reliability was low, relative to trials where visual reliability was high. Visually there appeared to be modulations in the relative beta and gamma power. To determine which (if any) of these effects were significant, we used cluster based permutation testing (see Methods).

Figure 4.2 shows the t-value (top) topographies, raw effect (middle) topographies, and power signal over time (bottom) resulting from the cluster analysis of the theta time-frequency representations. There were two clusters where relative auditory power scaled significantly with visual reliability during audio-visual trials. The first emerged in the theta band (averaged 4-6 Hz) over fronto-central electrodes, from stimulus onset to 150 ms (Figure 4.2A, cluster based permutation test, $t(19) = -73.7835$, $p = 0.011$). Specifically, at this epoch there was significantly lower theta power in the audio-visual condition where visual reliability was low, relative to the audio-visual condition where visual reliability was high. The second significant cluster emerged later in from 252 ms to 300 ms in the alpha band (averaged 10-12 Hz), again over fronto-central electrodes (Figure 4.2B, cluster permutation tests, $t(19) = 19.714$, $p = 0.021$). This time, there was significantly higher alpha power in the audio-visual condition where the visual reliability was low, relative to the condition where visual reliability was high. We found no significant beta or gamma activity, and therefore excluded it from further analysis (cluster randomisation tests $p>0.05$).

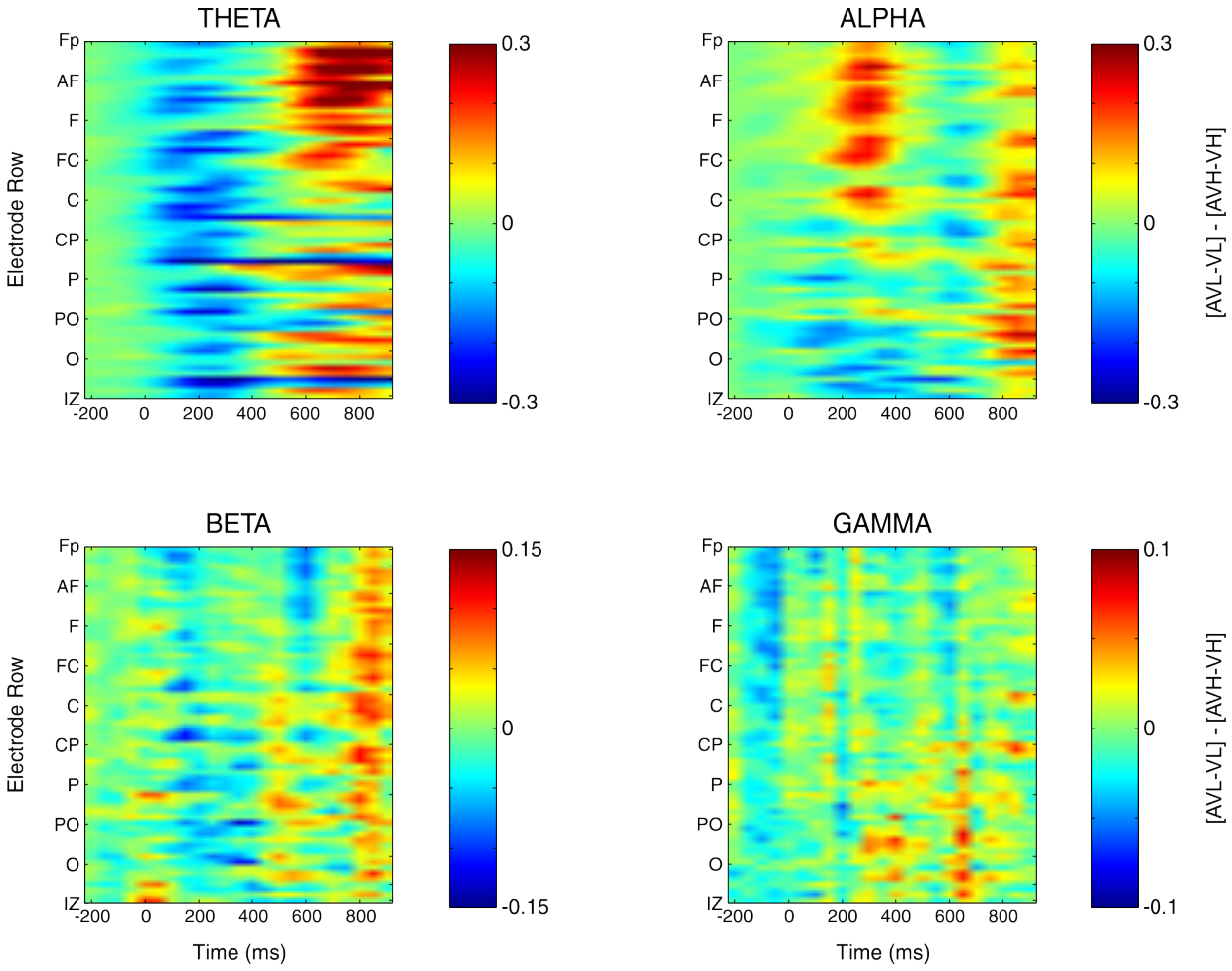


Figure 4.1 | Auditory Power Scaling with Visual Reliability. Each plot shows the grand average (n = 20) of the difference in auditory power elicited during audio-visual trials, relative to stimulus reliability. This is calculated as: [AVL - VL] - [AVH - VH]. For each individual plot, the data is averaged over the frequency band (see Methods for specification of bands). All plots shows the power difference spectra [AVL - VL] – [AVH - VH] for theta (top left), alpha (top right), beta (bottom left) and gamma (bottom right). Y axis in all graphs indicates arrangement by electrode row: frontal electrodes at top of y axis, posterior electrodes at bottom of y axis (see Supplementary Figure S4.2 for row positions). Positive values indicate there is higher (relative to baseline) relative power for the condition where visual reliability was low. Negative values indicate there is higher (relative to baseline) power for the condition where visual reliability was high. Colorbars represent the difference in baseline normalised power (relative power) (note different colorbar scales for each plot). Stimulus stream onset occurred at 0 and offset at 900 ms.

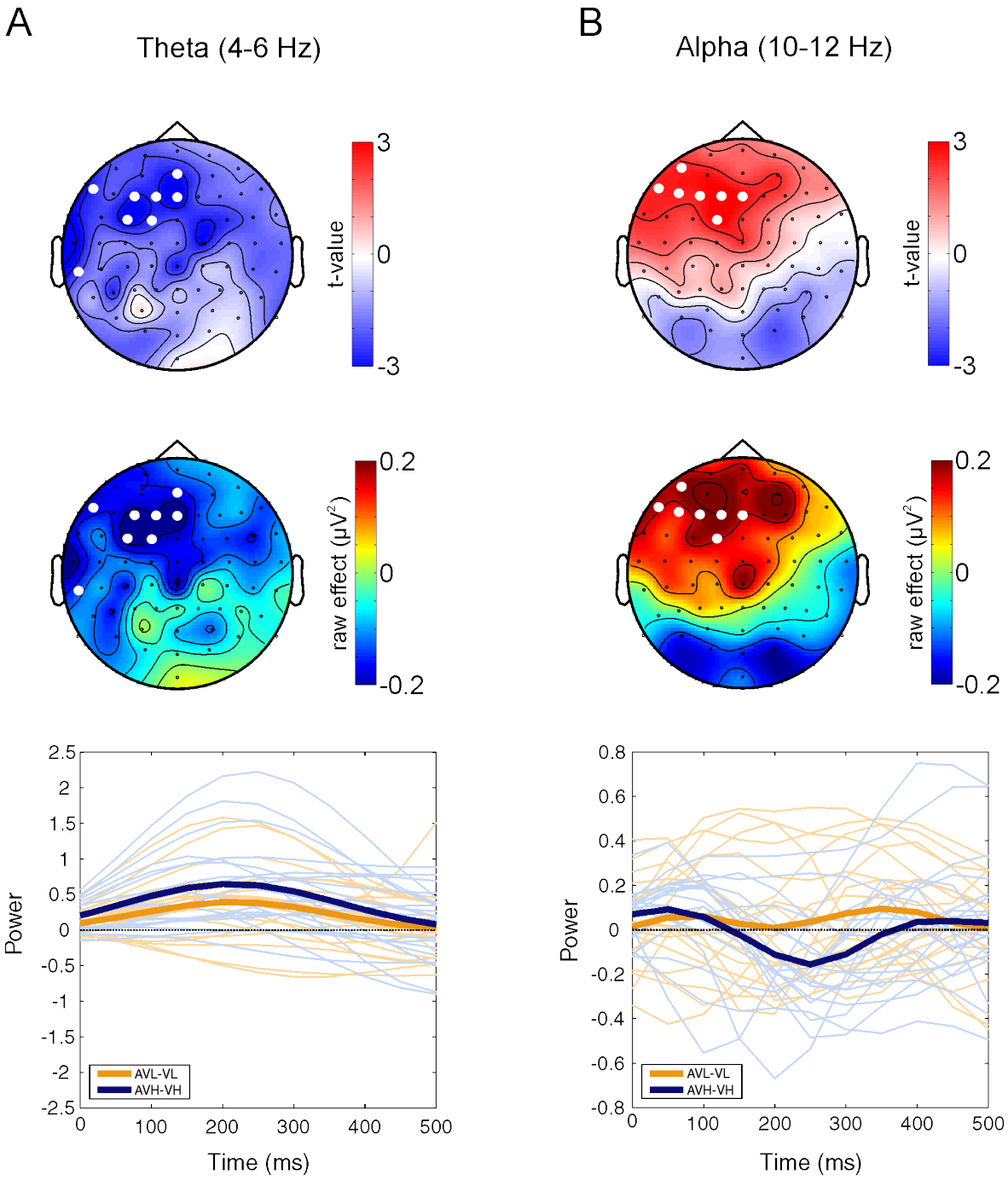


Figure 4.2 | Significant cluster differences. Cluster permutation tests revealed significant relative differences in the theta (left) and alpha (right) power related to auditory activity, depending on visual reliability ([AVL - VL] vs [AVH - VH]). (A) Scalp topography shows t-values (top) and raw effect (middle) for theta power averaged over the significant time window (0 to 150 ms). Electrodes showing significant differences are highlighted by white circles. For theta, these are: AFz, Fz, F1, F3, FC, FC3, F7, TP7. For alpha, these are: AF7, F7, F5, F3, F1, F2, FC1. (A) (bottom) shows the group averaged theta power for each reliability condition separately for an early time window during stimulus presentation. (B) Shows the same as in (A) for the alpha power (significant time window: 252 ms to 300 ms).

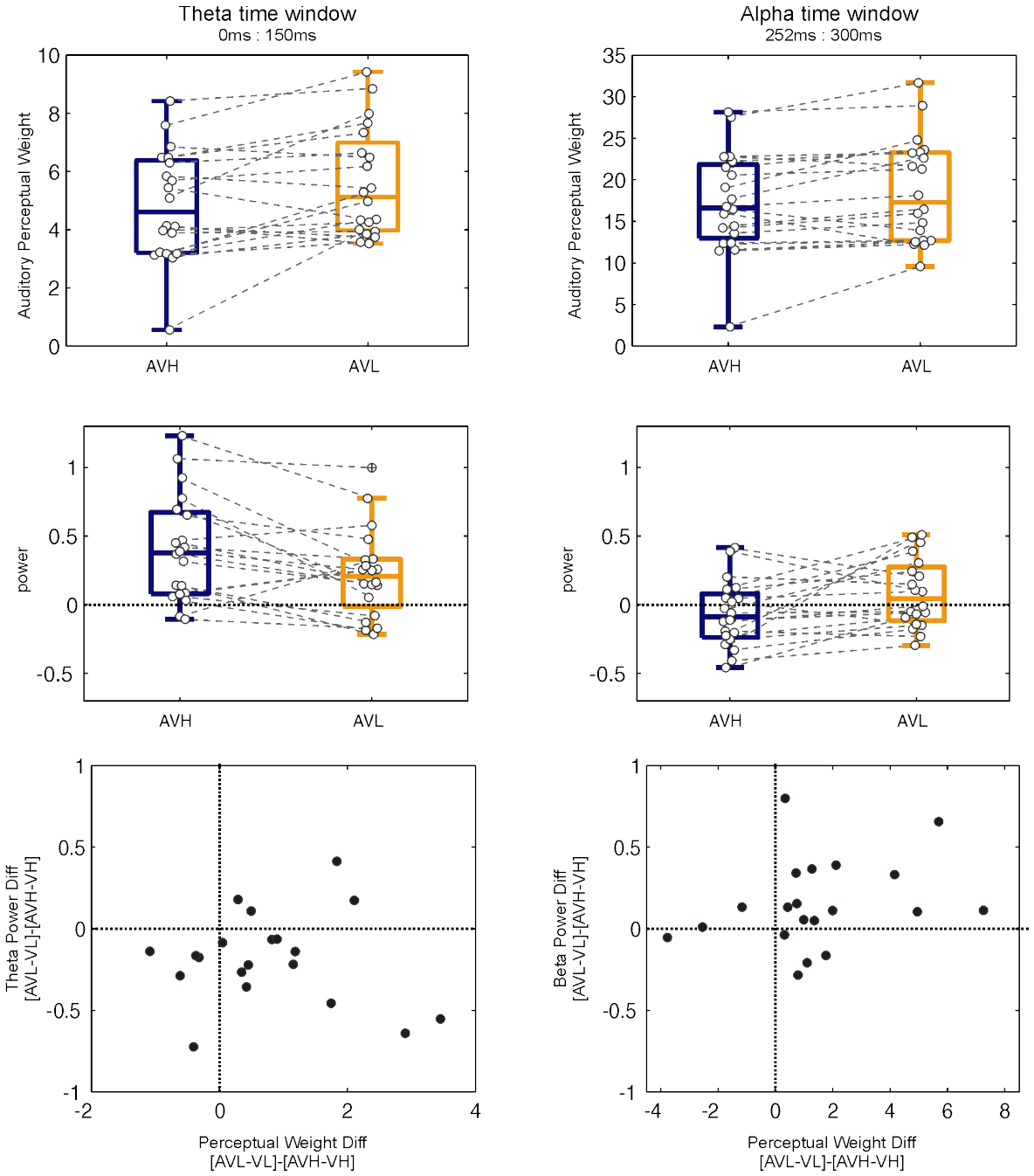


Figure 4.3 | Behavioural and Neural Relation. (A) Theta power was averaged across frequency (4-6Hz), time (0:150 ms) and electrodes (see Figure 4.2) of the significant cluster. (A, top) shows the decrease in auditory perceptual weights from high reliability (AVH) to low reliability (AVL) during the time window of significant theta power change (0 to 150 ms). Auditory perceptual weights were averaged across the time window for each subject individually. (A, middle) shows the theta power in audio-visual high (AVH) and audio-visual low (AVL) reliability conditions during the same time window. (A, bottom) shows the relation between these averaged perceptual weight and theta power values. Panel (B) shows the same as (A) for perceptual weights (top), alpha power (middle), and relationship between the two (bottom) the time window of significant alpha change (252 ms to 300 ms). Note: these plots contain the relative difference in auditory power between visual reliability conditions during audio-visual trials ([AVL - VL] - [AVH - VH]) (with power relative to baseline).

Figure 4.3 shows the perceptual weight change (top row), power change (middle row), and correlation between power and perceptual weights (bottom row) in each reliability condition (AVH and AVL), for theta (Figure 4.3A) and alpha (Figure 4.3B) separately. During the epoch theta power was modulated by visual reliability (0 to 150 ms) auditory perceptual weights significantly increased (median auditory weight AVH = 4.608, $t(19) = -2.889$, $p = 0.009$, effect size = -0.396).

During the epoch alpha power was significantly modulated by visual reliability (252 ms to 300 ms), again auditory perceptual weights significantly increased (median auditory weight AVH = 16.616, AVL = 17.291; $t(19) = -2.4644$, $p = 0.0234$, $g = -0.227$). However there was no significant correlation between power change and perceptual weight change for either theta or alpha (theta: $r = 0.03$, $p = 0.8$; alpha: $r = 0.23$, $p = 0.33$). Taken together, these results demonstrate that early in the trial (<300 ms), increases in auditory perceptual weights emerge at the same time as lower theta power and higher alpha power as a result of reduced visual reliability. However, the relationship between the two needs further investigation.

Discussion

In the present study we examined how oscillatory activity was modulated by the reliability of sensory information during an audio-visual rate discrimination task. Our results revealed lower theta power and higher alpha power emerged in conditions when visual reliability was low, relative to conditions when visual reliability was high. These modulations emerged early in the trial (from stimulus onset to 300 ms) over fronto-central electrodes, and occurred alongside a significant increase in auditory perceptual weights. Together these results suggest a role for theta and alpha in sensory weighting, and provide the second set of results in this thesis demonstrating early multisensory effects emerging from the brain during audio-visual perception.

As the behavioural aspects of sensory cue weighting have already been discussed in Chapter 3, the following section will concentrate on discussing the observed oscillatory power modulations and how they relate to cue reweighting.

Theta and Alpha: Audio-visual Processing

Theta and alpha oscillations have been widely found in previous experiments investigating audio-visual processing. For example, a study by Sakowitz, Quiroga, Schurmann, & Basar, (2005), found that strong, early (<100 ms) fronto-central theta oscillations, and later (>200 ms) alpha oscillations emerged over fronto-central regions during audio-visual trials, compared to unisensory auditory or visual trials. Based on the timing difference between the theta and alpha oscillations, the authors speculate that each frequency could underlie different stages of audio-visual perception in the brain. In our study, we found similar results as the Sakowitz et al,. (2005) study: in our work, theta and oscillatory activity emerged early and sequentially (<150 ms for theta, >250 ms for alpha) during audio-visual integration over fronto-central regions. Thus, our results provide support for their hypothesis that each frequency underlies a different stage of perception.

Additionally, as we observed modulations in oscillatory power due to sensory reliability, we can add to this hypothesis by speculating that theta and alpha each reflect different stages of the cue weighting process during perception. Specifically, we speculate that that the early theta response (<150 ms) could represent encoding of sensory reliability while the later alpha (>200 ms) represents correlates of perceptual reweighting, or indeed may reflect the process of cue reweighting in the brain. This also echoes the findings of the previous chapter, which revealed that the earliest neural correlates underpinned the reliability of the sensory information (84 ms), while later components (120 ms and 204 ms) were related to perceptual weighting. Taking these results into account lends strength to our interpretation. However, it must be stressed that these results are speculative, and given that the changes we observed in power were not correlated with changes in behaviour, the modulations could just reflect degraded stimulus processing rather than a correlate of behavioural reweighting. More work investigating the neural oscillations underlying cue weighting is needed to shed more light on this interpretation.

Theta and Alpha: Multisensory Attention and Inhibition

Theta and alpha have also been found in studies investigating multisensory attention. For example, a study by Keller, Payne, & Sekuler, (2017) examined oscillatory patterns during an audio-visual oddball task where subjects had to divide attention between the two streams. The results showed that fronto-central theta was stronger in the audio-visual divided attention condition compared to the non-divided condition. This finding was replicated in a follow up experiment (Keller et al., 2017). Together, these findings strongly suggested that theta plays a prominent role in attention to multisensory stimuli.

Alpha has also often been linked to attentional processes (Klimesch, Doppelmayr, Russegger, Pachinger, & Schwaiger, 1998; Klimesch, 2012; Rohenkohl & Nobre, 2011; Thut, Nietzel, Brandt, & Pascual-Leone, 2006). Specifically, studies have shown increased alpha activity in preparation of attention (Worden, Foxe, Wang, & Simpson, 2000) and as a result of directed attention early after stimulus onset (Yamagishi et al., 2003). Complementary findings of other studies have shown that an increase in alpha power occurs in response to attentional shifts away from visual stimuli (Sauseng et al., 2005; Thut et al., 2006; Yamagishi, Goda, Callan, Anderson, & Kawato, 2005). Together, these results show increased alpha accompanies directed attention towards one modality.

In terms of linking such effects to our results, we can interpret the lower theta observed in low visual reliability conditions as an indicator of less divided attention (as in Keller et al., 2017), and higher alpha power in low visual reliability conditions as a result of directed attention (as in Sauseng et al., 2015; Thut et al., 2006). This interpretation fits in well with our experimental design, as the initial lower theta power in conditions where auditory reliability was high and visual reliability was low could reflect the initial decreased need for divided attention as a result of less reliable visual stimulus. Similarly, the higher alpha power seen in the same conditions (high auditory reliability, low visual reliability) could reflect a shift towards increased focused attention on the more reliable auditory stimulus. Both of these modulations in power also coincide with a shift towards increased auditory weighting and decreased visual weighting as a

result of changing stimulus reliability. Together, this is one interpretation of our results that fits with the findings of the attention literature.

Related to this, alpha has also been implicated as an inhibition mechanism by which task-irrelevant or unattended stimuli are suppressed (Foxe & Snyder, 2011; Händel, Haarmeier, & Jensen, 2011; Jensen, Bonnefond, & VanRullen, 2012). For example, Payne et al., (2013) found consistent alpha power when subjects were instructed to ignore a visual stimulus, and this alpha power was a good predictor of accuracy recall for the auditory stimulus. They concluded that alpha power was therefore related to the suppression of the visual stimulus. In another study, Keller et al., (2017) showed alpha oscillations persisted during the presentation of auditory stimuli, but not visual. Consequently, they speculated this was because participants were suppressing the visual fixation cross, in order to divert attention to the auditory stream. Two other studies (Foxe, Simpson, & Ahlfors, 1998; Worden et al., 2000) found results that support this suggestion, revealing that alpha oscillations act as an anticipatory mechanism for suppressing distracting visual information. Fu et al., (2001) extended this work to a crossmodal task, and found significantly higher alpha emerged when subjects were cued to attend to the auditory modality. Again, this suggested that alpha oscillations may serve to suppress visual attention during multisensory integration. Finally, alpha has even been found to show a causal relationship with visual stimuli suppression. Romei, Gross, & Thut, (2010) used TMS to stimulate within the alpha frequency range, and showed directly that the visibility of visual targets was significantly impaired as a result. This did not occur with stimulation in the theta and beta band. This strongly suggests a causal role for alpha in visual suppression.

In terms of our results, alpha being associated to visual stimulus suppression fits with our above interpretation that theta and alpha activity represent the decreased need for divided attention and directing of attention towards the auditory modality when visual reliability is low. However, we would need to modify the interpretation slightly: instead of alpha power representing diverted attention away from the visual towards auditory, it may specifically represent suppression of attention to the visual stimulus, which results in the directed auditory attention. One could thus assume that the modulation in alpha power we see results from a brain process that suppresses the visual information when

it is of low reliability, in order to place greater emphasis on the more reliable auditory information and generate a more accurate perception.

However, while this interpretation of increased alpha being involved in attention and suppression fits nicely with our results, it is difficult to tease apart whether the modulations we see in the alpha band are related to attention directed towards the auditory stimuli, suppression of the visual stimuli, or both. Furthermore, the literature is complicated by the findings of other studies that show attention is associated with modulations in power in all frequency bands and that conclude attention cannot be related to any single frequency band (Fan et al., 2007). Thus, more work is required to dissociate the role of alpha in auditory attention and visual suppression.

Theta and Alpha: Memory

Finally, theta and alpha have been implicated in memory processes. In one study, Jensen & Tesche, (2002) asked subjects to retain lists of numbers of varying length (1,3,5 or 7 numbers), and found that frontal theta power increased as memory load increased. In another study, Meltzer et al., (2008) analysed oscillatory activity recorded from intracranial recordings during a memory task, and showed increases in theta and alpha occurring over midline frontal electrodes, as well as decreases in both bands occurring over occipital cortex, frontal and parietal regions in response to increasing memory load. Finally, in a combined fMRI-EEG study, Scheeringa et al., (2008) showed posterior alpha and frontal theta increased with increases in working memory load. Taken together, the findings of these studies (and others, e.g. Raghavachari et al., 2001; Roux & Uhlhaas, 2014) point towards a role of theta and alpha in working memory maintenance. Again these findings from the field tie in with our paradigm, as in our experiment, there was a memory component; subjects had to compare the stimulation rate of two sequentially presented stimuli streams. Thus, it could be that the observed modulations in theta and alpha oscillatory power relate to task related memory processes rather than sensory processing. For example, it may be that subjects find low reliability stimuli harder to remember than the salient, high reliability stimuli, and this is reflected in modulations in the power of oscillatory activity.

However, we feel that this interpretation is the least likely of the potential ones presented above, for two reasons. First, in terms of working memory load, subjects were not required to remember how many stimuli were presented in a trial, rather they were only required to judge which stream (out of two) flickered at a higher rate. Additionally, in our experiment there were easy trials (8 and 14 Hz compared to 11Hz) and hard trials (10 and 12 Hz compared to 11Hz) but they were all of equal length (900 ms). Thus, while there were differences in task difficulty, there were not differences in overall working memory capacity as the stimuli were all of equal length regardless of rate. Together these reasons make it probable that memory load did not play a role.

Secondly, our results showed significant differences evident from stimulus onset but only until 300 ms; if theta and alpha modulations were related to maintaining a representation of the stimulus rate, and these varied with reliability, we expect we would see modulations in power emerge throughout the entire trial (which was 900 ms long). As this was not the case, it suggests that the significant modulations in theta and alpha seem are not due to memory processes. However, future work on this dataset could explore the neural oscillations underlying each rate presented to examine for differences based on increasing number of stimuli to be retained. While beyond the scope of this chapter, this proposed analysis provides an interesting starting point for future work.

Overall, the broad range of processes to which theta and alpha have been linked makes it difficult to conclusively assign a functional process to our results. In addition, there are different aspects of the time-frequency representation still to be investigated here, such as phase and cross frequency coupling. Future work could build on the findings here by such analyses on the dataset to examine the underlying effects in more detail. Yet, in line with the literature above, our speculation is that the relative lower theta power and higher alpha power we observed during conditions where visual reliability was low is tied to either sensory reweighting processes or attention, or a combination of both.

Chapter 4: Summary

This chapter has presented results demonstrating that oscillatory activity is also modulated by sensory reliability at early points during a trial, and that modulations in oscillatory patterns occur simultaneously alongside behaviour. As in Chapter 3, this was achieved using single-trial analysis and the attempt to link neural signals to behaviour, rather than demonstrating a simple modulation of evoked response amplitudes in response to multisensory stimulation. Overall, the work presented in this Chapter provides the first insights into the oscillatory patterns underlying sensory cue weighting during audio-visual integration, and the results tie in with the existing literature as well as the results found in Chapter 3. This chapter also presents the second set of results showing early modulation of neural signals during multisensory perception, once again lending support to the modern view that the entire cortex is capable of multisensory processing.

Supplementary Materials: Chapter 4

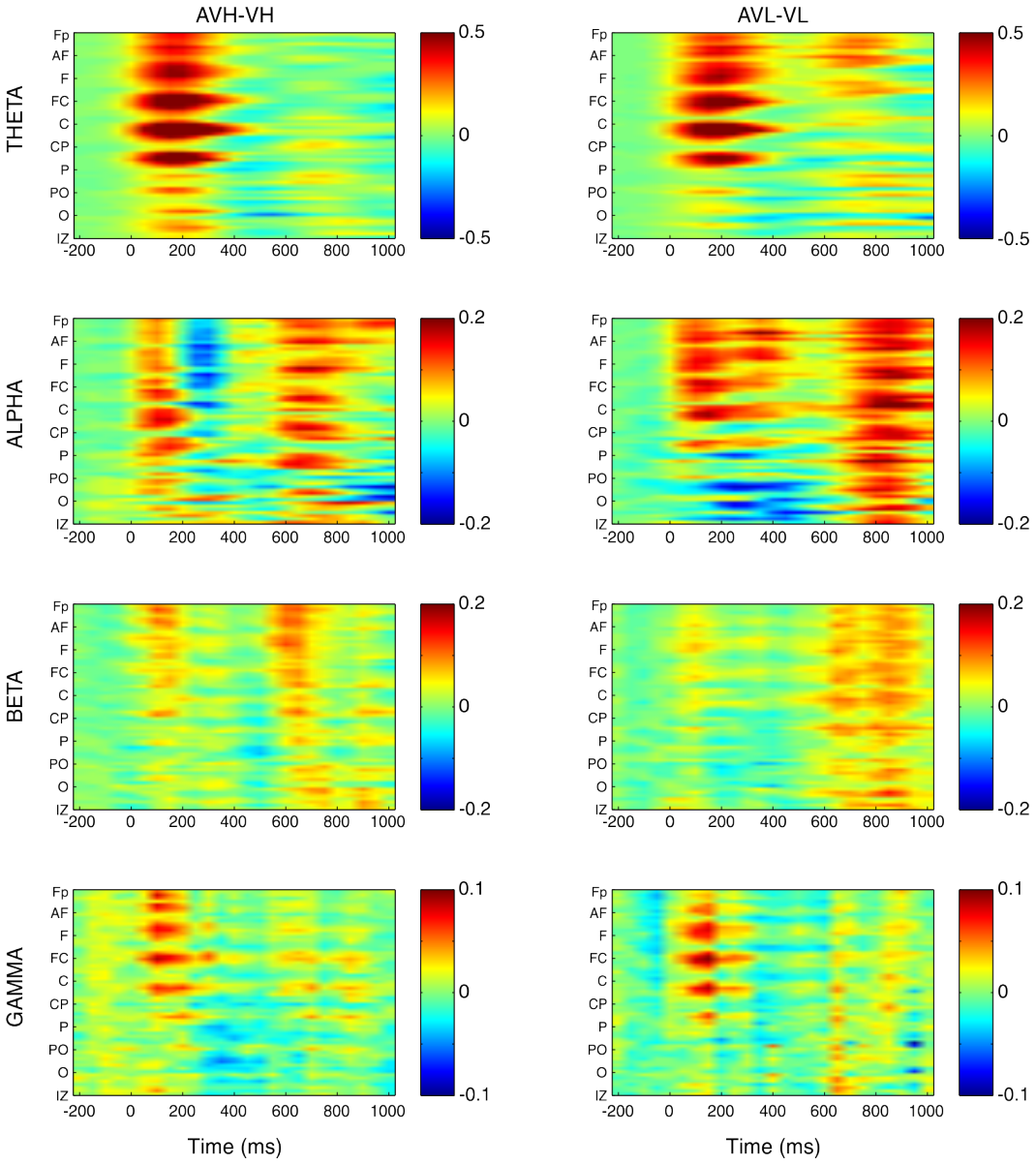


Figure S4.1 | Auditory Power Scaling with Visual Reliability for Individual Reliability Conditions. Each plot shows the grand average (n = 20) of the difference in auditory power elicited during audio-visual trials, relative to stimulus reliability. For each individual plot, the data is averaged over the frequency band (see Methods for specification of bands). Y axis in all graphs indicates arrangement by electrode row: frontal electrodes at top of y axis, posterior electrodes at bottom of y axis (see Supplementary Figure S4.2 for row positions). Colorbars represent relative power (baseline-normalised power).

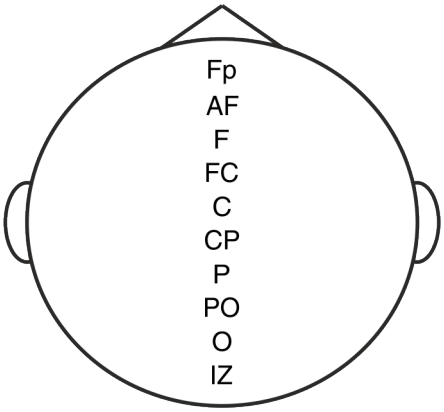


Figure S4.2 | Electrode Rows. All TFR plot Y axes are arranged according to the row layout above.

Chapter 5 : Neural Correlates of an Auditory Pitch - Visual Size Cross-modal Association Emerge Early during Perception

Introduction

While the brain predominately associates sensory cues based on their spatial or temporal congruency, it also embodies some more implicit cross-modal associations. For example, research has shown that human subjects often associate high pitched tones with small objects, and low pitched tones with large objects (Bien et al., 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Mondloch & Maurer, 2004; Parise & Spence, 2008; Parise & Spence, 2009, 2012). Similarly, high pitch tones are often paired with high elevations in space, while low pitched tones are paired with low elevations (Evans & Treisman, 2010; Melara & O'Brien, 1987; Roffler & Butler, 1968; Widmann, Kujala, Tervaniemi, Kujala, & Schröger, 2004). Although these two pairings are the most consistently demonstrated, many other associations have been found, including: brightness-loudness (bright object-loud sound vs. dark object-quiet sound, Stevens & Marks, 1965); pitch-brightness (high pitch-bright object vs. low pitch-dark object, (Mondloch & Maurer, 2004); pitch-angle (high pitch-sharp angle vs. low pitch-obtuse angle, Parise & Spence, 2012); and pitch-spatial frequency (high pitch-high spatial frequency vs. low pitch-low spatial frequency, Evans & Treisman, 2010; Parise & Spence, 2012). Collectively, these interactions are known as "cross-modal associations", with preferred pairings defined as "congruent" and non-preferred pairings defined as "incongruent".

Over the years, cross-modal associations have been examined in a variety of ways. Early studies used explicit paradigms, where subjects directly matched stimuli based on their preferred congruency (Kohler, 1929; Ramachandran & Hubbard, 2001; Stevens & Marks, 1965). In more recent years, studies focused on examining how stimulus congruency modulates behaviour using implicit paradigms. For example, spatial and temporal ventriloquist paradigms have been used to examine how cross-modal associations modulate task performance. In such paradigms, subjects are presented with two stimuli (one auditory, one visual) at either a spatial or temporal offset, and asked to judge the location or

timing of both stimuli. Associations between stimuli are then measured by examining the extent to which the perceived location of an auditory stimulus is affected by a spatially disparate visual stimulus (spatial ventriloquism), or the extent to which the perceived temporal rate or order of a visual stimulus is affected by a temporally disparate auditory stimulus (temporal ventriloquism). For example, Bien et al. (2012) demonstrated that the reported location of the auditory stimulus was "pulled" towards the visual stimulus location to a greater extent when the stimuli were congruent (high tone-small circle, low tone-large circle) compared to incongruent (high tone-large circle, low tone-small circle). Similarly, Parise & Spence (2008) used a temporal ventriloquist paradigm to show that subjects were better able to judge the temporal order in which two visual stimuli (small and large circle) had been presented, when the stimuli were paired with congruent auditory stimuli (high and low tones respectively), compared to incongruent auditory stimuli (low and high tones respectively).

Reaction time can also be modulated by cross-modal congruency, as demonstrated by speeded classification paradigms. Such paradigms present subjects with two stimuli, and ask them to make a judgement about one stimulus as quickly and accurately as possible. Following the assumption that the simultaneous presentation of the stimuli provides behavioural facilitation (even if one stimulus is task irrelevant), compatible stimulus pairings should lead to shorter reaction times compared to incompatible pairings. Indeed, many studies have used these paradigms to present subjects with congruent or incongruent pairings of stimuli (as defined above), and have demonstrated that congruent pairings lead to shorter reaction times than incongruent pairings across a variety of stimulus combinations (pitch-size, Gallace & Spence, (2006); pitch-spatial position, Evans & Treisman, (2010); Rusconi et al., (2006); and pitch-spatial frequency, Evans & Treisman, (2010)).

Interestingly, cross-modal congruency can modulate reaction times even when only one stimulus is presented on a trial. In a recent study, Parise & Spence (2012) used a version of the Implicit Association Test (IAT, Greenwald, McGhee & Schwartz, 1998) to test for cross-modal associations. The IAT is a computer task designed to measure associations between stimuli via stimulus-response key mapping. The main assumption of the IAT is that if an association exists between two stimuli, response times should be faster when these associated stimuli are

assigned to the same response key as compared to when they are assigned to different response keys (Greenwald, McGhee, & Schwartz, 1998). Importantly, this modulation of reaction times should occur even when only a single stimulus is presented per trial. Using this measure, Parise & Spence (2012) demonstrated that response times were faster when congruent pairings of stimuli were assigned to the same response key, compared to when incongruent pairings were, across a variety of audio-visual stimulus combinations. Given that these findings were observed even when only one stimulus was presented per trial, and when no explicit reference to the pairings of the stimuli were involved in the task, the results suggest that cross-modal associations emerge implicitly at an early stage of perception, rather than a later, explicit decision related stage. However, what stage or perception these associations emerge at is still a source of much debate (for review, see Spence & Deroy, 2013).

One way to shed light on this issue of when cross-modal associations emerge is to examine the underlying neural signals. One study by Bien et al., (2012) used electroencephalography (EEG), transcranial magnetic stimulation (TMS), and a spatial localisation task to examine an auditory pitch-visual size association. They found that parietal and frontal ERP components were modulated by cross-modal congruency early in the trial (around 250-300 ms), and demonstrated that applying TMS over parietal cortex — an area known for multisensory processing (Matsuhashi et al., 2004; Molholm, 2006; Pasalar, Ro, & Beauchamp, 2010) — could abolish the association. Similarly, Kovic, Plunkett, & Westermann, (2010) found early ERP effects at occipital (~140 ms) and parietal (~340 ms) sites were sensitive to the learned congruency of an association between semantic (auditory) words and visual objects. Finally, Sadaghiani, Maier, & Noppeney, (2009) used fMRI to demonstrate that associations between auditory and visual motion signals emerged in motion areas, whereas higher-level speech-motion associations emerged in fronto-parietal areas. The conclusions of each study generally converged: Bien et al., (2012) concluded their results depended on perceptual processes due to the early onset and abolishment of the effect via TMS over a known multisensory perceptual area. Kovic et al., (2010) concluded similarly, given the early onset of their results (i.e. during stimulus presentation). Sadaghiani et al., (2009) suggested that the origin of the effects depends on the stimulus: both perceptual level and higher-level associations

were demonstrated across the cortex but depended on the whether the association was between basic stimulus features, or a higher level semantic association. Overall, the conclusions made by the authors of these studies generally supported the early, perceptual origin of cross-modal association.

However, these studies were limited in their ability to accurately determine the origin and timing of cross-modal associations, as they relied on few recording sites (Kovic et al., 2010), or sampled brain activity at low temporal resolution (Sadaghiani et al., 2009). Additionally, the studies relied on paradigms which required explicit matching, or where two stimuli are presented simultaneously (Bien et al., 2012). As a result, it remains unclear whether the behavioural effects seen were due to genuine cross-modal associations, or to confounding effects related to dual stimulus presentation (e.g. behavioural facilitation due simply to the presence of two stimuli) or decreased response time as a result of divided attention. Together, these factors make it difficult to draw clear conclusions about where and when brain activity reflects true cross-modal associations.

In this study we addressed these questions by examining when effects of an auditory pitch – visual size cross-modal association emerged in neural signals. To do so, we used the modified version of the implicit association test (IAT) as in Parise & Spence (2012), combined with EEG based neuroimaging and two approaches to single-trial EEG analysis. Importantly, using the IAT overcomes all the methodological issues mentioned above: it presents only one stimulus per trial, thus avoiding attentional or multisensory confounds, and it manipulates congruency by changing the stimulus-response key mapping across blocks, thus avoiding explicit matching and subjective reports. Additionally, the presentation of only a single stimulus on each trial allows us to extract sensory-specific processes from brain activity and relate these to behaviour on a trial-by-trial basis.

In line with previous IAT studies, we hypothesised that reaction times would be faster when congruent pairs of stimuli (high tone-small circle, and low tone-large circle) were assigned to the same response key, compared to when incongruent pairs were (low tone-small circle, and high tone-large circle). However, in our study we restricted the stimulus-response manipulation to

auditory trials only, and held the assignment of the visual stimuli constant (see Methods). As a result, we hypothesised reaction times would be shorter for congruent auditory trials, but not congruent visual trials. We did not expect performance to be significantly modulated by congruency for two reasons: first, past work has typically shown no modulation of performance due to congruency (Parise & Spence, 2012). Second, the task was a simple, single stimulus identification task where there were only four repeating options, and so subjects could be expected to perform well. We had two hypotheses regarding the EEG data: a) that brain activity sensitive to the task-relevant sensory feature (e.g. acoustic pitch or visual size) should be modulated by congruency, and b) that these neural correlates sensitive to the cross-modal congruency should be predictive of subject's single trial reaction times. Finally, we had no prior hypothesis as to when these effects would manifest during a trial or where they would localize. However, if implicit associations arise at an early perceptual level, one would expect their neural correlates to emerge with short latencies after stimulus onset, likely over early sensory areas. In contrast, if these associations arise at a decisional level, one would expect neural correlates to arise only later and potentially over associative and higher-order brain regions.

Methods

Subjects

20 participants (13 females; age range 19-32) took part in the study. One subject's data (S19) had to be discarded due to noisy EEG channels. All subjects reported normal or corrected to normal vision and normal hearing. Subjects were recruited via the University of Glasgow Subject Pool, and received £6 per hour for their participation. The study was approved by the local ethics committee (application number: 300130001, College of Science and Engineering, University of Glasgow) and conducted in accordance with the Declaration of Helsinki.

Stimuli

Stimuli were created and presented using MATLAB (MathWorks) and the Psychophysics Toolbox Extensions (Brainard, 1997). Visual stimuli consisted of

two light grey circles ('small' and 'large', 2cm and 5cm, 1.1 ° and 2.8 ° of visual angle respectively) presented for 300 ms atop a darker grey background (Figure 5.1). Auditory stimuli consisted of two 300 ms pure tones ('high' and 'low' pitch, 2000Hz and 100Hz respectively). The sound intensity of each tone was matched to 72 Db_A SPL (left and right ear) using a sound level meter. Auditory stimuli were presented using Sennheiser headphones and visual stimuli were presented on a Hansol 2100A CRT monitor at a refresh rate of 85 Hz.

Task

The task was a modified version of the IAT (Greenwald, McGhee, & Schwartz, 1998b), as used in Parise & Spence, (2012). As mentioned previously, the IAT measures implicit associations via manipulating stimulus-response key mapping, and assumes response times will be faster when associated pairs of stimuli are assigned to the same response key. In this modified version, on each block, one auditory and one visual stimulus were assigned to the left response key, and one auditory and one visual stimulus were assigned to the right (two stimuli per key, see Congruency section for assignment). Subjects were then presented with one stimulus per trial, and asked to identify which was presented as quickly and accurately as possible using the appropriate response keys. Implicit associations were measured via reaction time modulations.

Congruency

Congruency was manipulated by changing the stimulus-response key pairings across blocks (Figure 5.1). On congruent blocks the small circle and high tone were assigned to the left response key and the large circle and low tone were assigned to the right key On incongruent blocks the small circle and low tone were assigned to the left response key and the large circle and high tone were assigned to the right response key. Importantly, the auditory assignment changed across blocks, while the visual assignment always remained fixed. In total, subjects completed 8 blocks (4 congruent and 4 incongruent presented in a randomised order) for a total of 1280 trials (160 trials per block, 40 trials for each stimulus type).

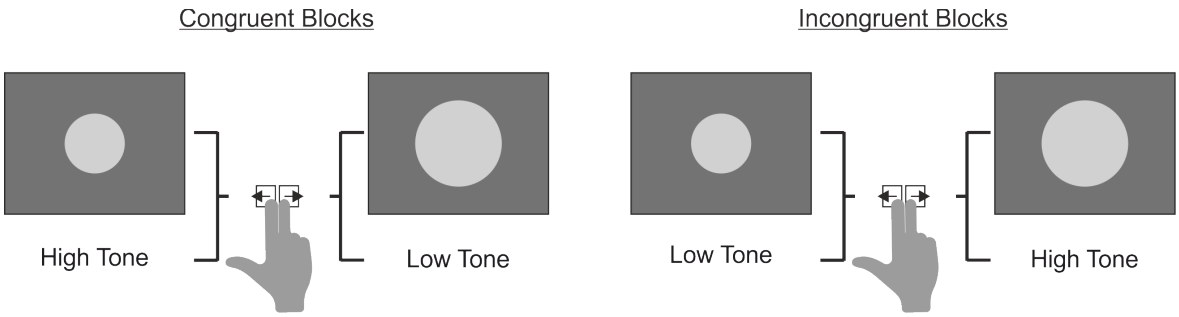


Figure 5.1 | Task. Subjects were presented with one stimulus on each trial and had to indicate which stimulus was presented as quickly and as accurately as possible using the left and right response keys. On each block, the assignment of visual and auditory features to the two response keys was manipulated (*left*, congruent pairings; *right*, incongruent pairings).

Procedure

The experiment was carried out in a dark and electrically shielded room. Each block began with instructions on the mapping between stimuli and response keys (see Congruency). Subjects were given as much time as they needed to memorise the instructions for the upcoming block. Each trial started with a fixation cross presented centrally for a randomised period (uniform distribution in 500 to 1000 ms). Then one of the four stimuli (see Stimuli) was selected randomly, and presented for 300 ms (Figure 5.1). Subjects had to respond as quickly as possible using the left and right keyboard keys, as defined by the block instructions (see Congruency). Subjects always responded using their right hand. Feedback was provided after each trial (green fixation cross for correct answers, red fixation cross for incorrect answers) for a randomised duration (uniform distribution from 300 ms to 600 ms).

EEG Recording and Preprocessing

EEG data was recorded using a 128-channel BioSemi system and ActiView recording software (Biosemi, Amsterdam, Netherlands). Signals were digitised at 512 Hz and band-pass filtered online between 0.16 and 100 Hz. Signals originating from ocular muscles were recorded from four additional electrodes placed below and at the outer canthi of each eye.

Data from individual subject blocks were preprocessed in MATLAB using the FieldTrip toolbox (Oostenveld et al., 2011) and custom scripts. Epochs of 2

seconds (-0.5 to 1.5 seconds relative to stimulus onset) were extracted and filtered between 0.5 and 90 Hz (Butterworth filter) and down-sampled to 200 Hz. Potential signal artefacts were removed using independent component analysis (ICA) as implemented in the FieldTrip toolbox (Oostenveld et al., 2011), and components related to typical eye blink activity or noisy electrode channels were removed. Horizontal, vertical and radial EOG signals were computed using established procedures (see Chapter 2, Preprocessing section and Hipp & Siegel, 2013; Keren, Yuval-Greenberg, & Deouell, 2010) and trials where there was a high correlation between eye movements and components in the EEG data were removed. Remaining trials with amplitudes exceeding ±120 μV were also removed. Successful cleaning was verified by visual inspection of single trials.

Analysis Methods

Analysis of Behavioural Data

For each subject, overall performance (proportion correct) and median reaction time (RT) were calculated separately for each modality and congruency. Trials with very fast (<300 ms) or slow (>1200 ms) responses were excluded (in line with Parise & Spence, 2012). Both RT and performance scores were submitted to Wilcoxon Signed rank tests for analysis (see Statistics). All reported RTs are calculated with respect to stimulus onset.

Analysis of EEG: Linear Discriminant Analysis

We used single-trial, multivariate linear discriminant analysis (Parra et al., 2005; Philiastides et al., 2014; Sajda et al., 2009) to extract discriminant components related to stimulus type, for each modality separately (high versus low pitch; small versus big circle). Prior to analysis, the data was band-pass filtered between 1 and 30 Hz. Classification was based on regularised Fisher's linear discriminant analysis (Philiastides et al., 2014), and was applied to the EEG activity in sliding time windows of 30 ms, at each 5 ms time point in the window from -300 ms pre-stimulus onset to 1 second post stimulus onset. The discriminant output (Y) was always aligned to the onset of the 30 ms window. Classification performance (Az) was determined using the receiver operator characteristic (ROC) and 10 fold cross-validation along with randomisation

testing (see Statistics). Scalp topographies showing the normalised correlation between the discriminant output and the EEG activity were estimated via a forward model (Philiastides et al., 2014).

To assess how these neural correlates of sensory evidence were modulated by congruency, the discriminant output (Y) for congruent trials was compared to that for incongruent trials using a cluster randomisation technique (see Statistics). We used linear regression to investigate whether the information contained in the discriminant component (Y) was predictive of behavioural reaction times. As we were interested in whether the quality of the sensory information reflected by the EEG component (i.e. the distance from zero, regardless of sign) was predictive of auditory reaction times, the discriminant output (Y) was flipped for trials that had been assigned to a negative value during classification (e.g. stimulus labels assigned to 0 were multiplied by -1). The discriminant output (Y) for each trial was then regressed against individual subject reaction times at each time point during the trial. Significance levels for both congruency and regression analyses were calculated using a cluster randomisation technique (see Statistics).

Analysis of EEG: Mutual Information

To assess how robust our effects were, we performed a complementary analysis based on mutual information (MI) (Gross et al., 2013; Ince, Jaworska, et al., 2016; Keitel et al., 2017). MI can be thought of as a likelihood ratio test for dependence between two variables of interest (e.g. between stimulus type and EEG, or between EEG and reaction times). A particular advantage of MI is that provides a common meaningful effect size scale (bits) across a wide range of statistical tests (Ince, Giordano, et al., 2016).

We calculated MI using a semi-parametric estimator: Gaussian Copula Mutual Information (GCMI) (Ince, Giordano, et al., 2016). This provides a data-efficient and robust lower bound approximation to MI by modelling the dependence between the variables with a Gaussian copula. However, no assumption is made on the marginal distributions of the variables. To allow direct comparison of the results obtained from MI analysis and the linear discriminant analysis, we follow the same pre-processing steps described in the previous section. EEG data were

band-pass filtered between 1 and 30 Hz, and then averaged in sliding windows of 30 ms from stimulus onset to 545 ms post stimulus onset in steps of 5 ms (with the data aligned to the onset of the 30 ms window). For each sensor we calculated the single trial central finite difference temporal derivative of the filtered EEG signal. At each time point of each trial, we added this temporal derivate to the voltage to obtain a bivariate response. GCMI allows us to estimate MI using this bivariate response. When a biphasic evoked potential is modulated by an experimental condition, considering voltage alone can often result in a double peak statistical effect, because of the zero crossing where the modulation changes sign (e.g. at the zero crossing of an amplitude modulated bi-phasic waveform). Since MI is an unsigned effect size this results in a two positive peaks, separated by a time period in which there is no significant effect. Including the EEG voltage and its instantaneous rate of change at each time point addresses this, since at the zero crossing ongoing modulation of the signal can be detected in the gradient. This gives a more balanced picture of the time window over which the EEG signal is modulated by the experimental condition (Ince, Giordano, et al., 2016). GCMI was computed between stimulus type (high/low tone and small/large circle) and the 2D EEG voltage response (EEG data, temporal derivative) for each modality, time point and electrode separately. Statistical significance was determined using randomisation analysis (see Statistics).

To examine when this MI statistic was affected by cross-modal congruency, for each modality we compared MI values (about high/low tone and small/large circle) computed from the congruent trials to MI values obtained from the incongruent trials at each time point. Significant clusters were determined using a cluster randomisation procedure (see Statistics). To examine the relationship between neural activity and behavioural responses, the EEG data in these significant clusters was then averaged over all significant electrodes and a 30 ms epoch over the centre cluster peak. This shorter epoch was chosen as averaging EEG activity over long (>100 ms) windows can cause problems with biphasic evoked potentials due to cancellation of positive and negative periods of the signal. Based on our filtering parameters, we chose a 30ms window around the peak to reduce noise and avoid including periods of signal with different sign.

The resulting single-trial EEG data in these clusters was then regressed against the single-trial reaction times for each participant using multiple linear regression (with each cluster as a predictor). Significance was determined by applying a permutation based t-test across subjects on the group level regression weights for each time point, and significant clusters determined using a cluster randomisation technique (see Statistics).

Statistics

All Z values reported were generated from a two-sided Wilcoxon signed rank test after testing assumptions of normality (which did not hold). Effect sizes were calculated by dividing the Z value by the square root of N (where N = the number of observations rather than subjects, Rosenthal, 1994). P-values were checked for inconsistencies using the R software package "statcheck" (Nuijten et al., 2016).

Statistical significance of classification performance (Az) were determined by randomly shuffling the condition allocated to each trial 2000 times, computing the group averaged Az value (area under ROC curve, see Methods) for each randomisation, and taking the maximal Az value over time for each randomisation. This built a distribution of Az values from which we extracted the 99^{th} percentile. Because of the maximum operation, this provides a Family-Wise Error Rate (FWER) of $p = 0.01$, corrected for the multiple comparisons over time points (Holmes, Blair, Watson, & Ford, 1996; Nichols & Holmes, 2001).

Significance for group-level effects of congruency on the discriminant output (Y) were obtained by comparing congruent and incongruent trials across subjects at each time point using a permutation based paired t-test across subjects (shuffled subject labels, 1000 permutations). Significant clusters were then determined using a cluster randomisation technique which compared the true t-value (resulting from comparing congruent to incongruent Y signals using true subject labels) to the shuffled t-value, based on a cluster threshold of $t = 1.8$ ($p<0.05$), maxsum cluster forming, minimum cluster size of 2, and cluster p-value = 0.05. Effect sizes were indicated as the equivalent r value that is bounded between 0 and 1 (Rosenthal & Rubin, 2003).

Significance levels for the LDA regression analysis were generated by randomly shuffling the trial specific reaction times and performing the regression analysis between the decoding signal and reaction times 1000 times. Significance was determined by applying a t-test against zero across subjects on the group level regression weights for each time point, and significant clusters determined using a cluster randomisation technique, with cluster thresholds set as described above.

Finally, statistical significance for the MI analysis was calculated using a randomisation test together with the method of maximum statistics (Holmes et al., 1996). For each time point, sensor, condition, and subject, GCMI was calculated 1000 times with permuted stimulus class labels. The maximum MI value over electrodes and time across permutations was calculated, and the 99th percentile used as the threshold for significance. MI values computed from congruent and incongruent trials were then compared at each electrode and time point using the same cluster randomisation technique (and settings) described above.

Results

Behavioural Results

Figure 5.2 shows the behavioural results. In line with our hypothesis, median reaction times were shorter for congruent versus incongruent trials (auditory congruent to incongruent: 959 ms to 993 ms, visual congruent to incongruent: 898 ms to 929 ms, calculated from stimulus offset). As expected, this difference was significant only for auditory stimuli (Wilcoxon sign rank tests: auditory, $Z = -2.1328$, $p = 0.033$, effect size = -0.3372; visual, $Z = -1.14487$, $p = 0.127$, effect size = -0.229). Performance score did not significantly differ between incongruent and congruent trials (Wilcoxon sign rank tests: auditory, median change congruent to incongruent, 94.3% to 94.1%, $Z = -0.402$, $p = 0.688$, effect size = -0.064; visual, median change congruent to incongruent, 97.4 % to 97.4%, $Z = 0$, $p = 1$, effect size = 0.002).

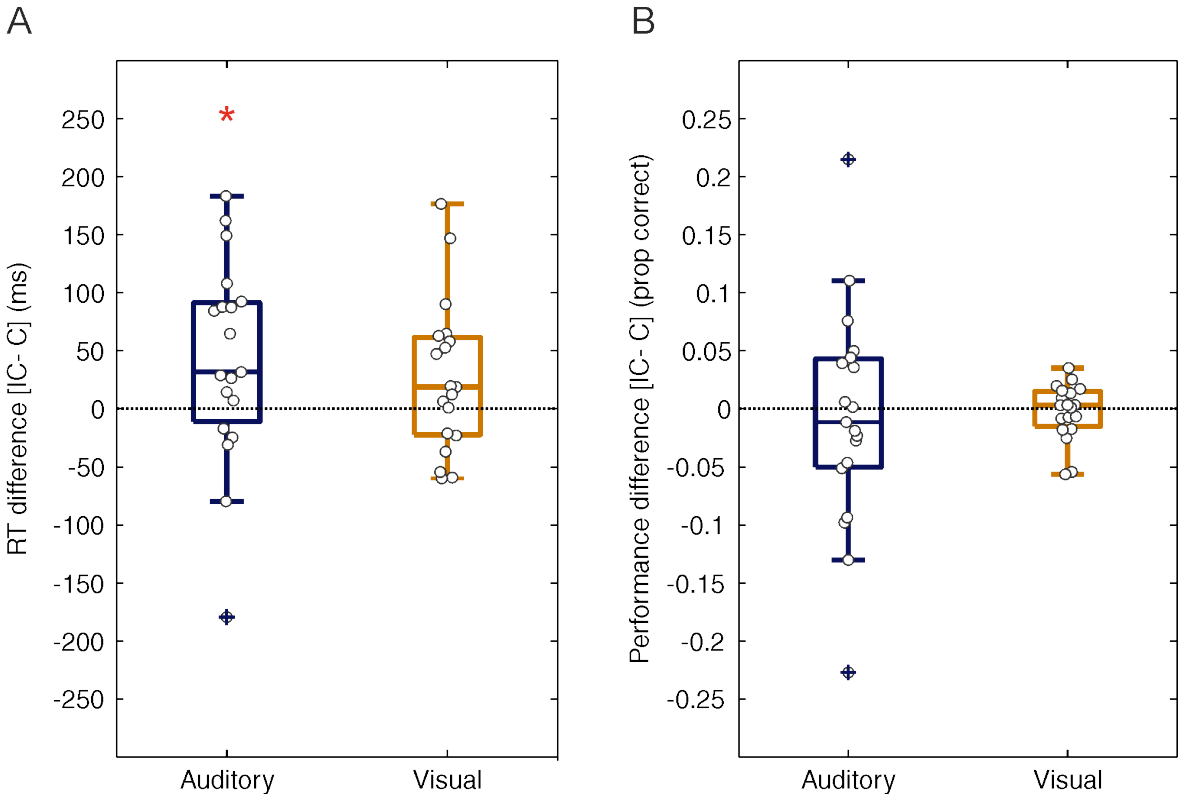


Figure 5.2 | Behavioural Results. (A) Median reaction time difference (Incongruent RT – Congruent RT) across all trials shown for each participant (grey circles). (B) Accuracy Difference (Incongruent – Congruent proportion correct). Asterisk (*) represents significant difference ($p<0.05$, Wilcoxon Signed Rank test).

EEG Decoding

Figure 5.3A displays the discriminant performance for auditory and visual trials. For auditory stimuli, significant performance emerged between 25ms after stimulus onset and 535 ms (Figure 5.3A, blue horizontal dotted line, cluster permutation test, $p<0.01$ Az value = 0.5387). The corresponding scalp models obtained from the correlation between the discriminant output and the EEG data (averaged over a 30 ms time window centred on the two classification performance peaks) revealed the strongest effects originated over posterior, central and temporal electrodes for both peaks (Figure 5.3A, topographical inserts, top row). These two topographies were very similar but of opposite sign. For visual stimuli, significant decoding performance emerged between 40ms after stimulus onset and 660 ms (Figure 5.3A, yellow horizontal dotted line, cluster permutation test, $p<0.01$ Az value = 0.5432). The corresponding scalp models (Figure 5.3A, topographical inserts, bottom row) showed strongest

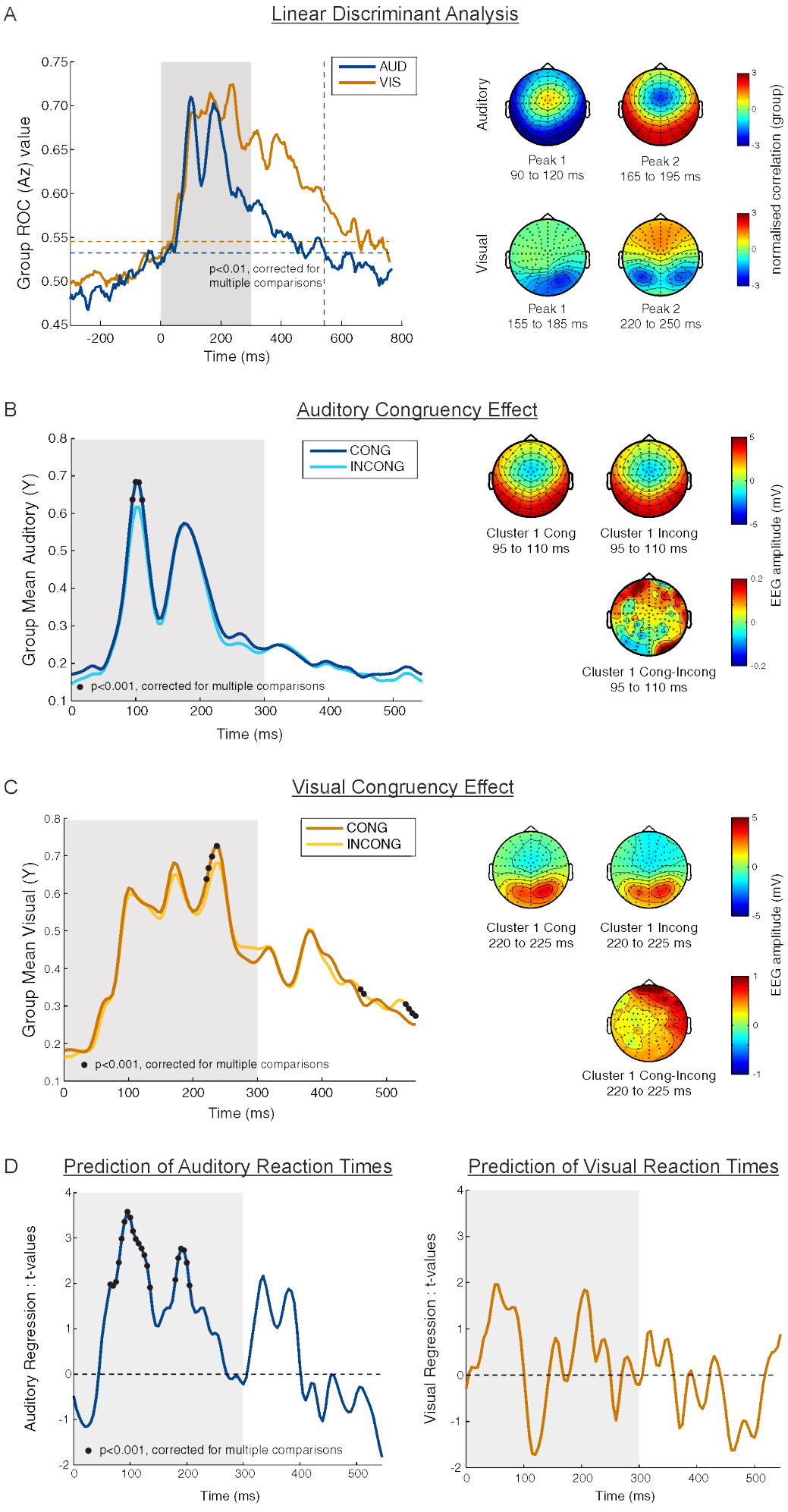

A
Linear Discriminant Analysis
AUD
VIS
Group ROC (Az) value
Time (ms)
p<0.01, corrected for multiple comparisons
Auditory
Visual
Peak 1
90 to 120 ms
Peak 2
165 to 195 ms
Peak 1
155 to 185 ms
Peak 2
220 to 250 ms
normalised correlation (group)
B
Auditory Congruency Effect
CONG
INCONG
Group Mean Auditory (Y)
p<0.001, corrected for multiple comparisons
Cluster 1 Cong
95 to 110 ms
Cluster 1 Incong
95 to 110 ms
Cluster 1 Cong-Incong
95 to 110 ms
EEG amplitude (mV)
C
Visual Congruency Effect
CONG
INCONG
Group Mean Visual (Y)
p<0.001, corrected for multiple comparisons
Cluster 1 Cong
220 to 225 ms
Cluster 1 Incong
220 to 225 ms
Cluster 1 Cong-Incong
220 to 225 ms
EEG amplitude (mV)
D
Prediction of Auditory Reaction Times
Auditory Regression : t-values
p<0.001, corrected for multiple comparisons
Prediction of Visual Reaction Times
Visual Regression : t-values
Time (ms)

Figure 5.3 | EEG Linear Discriminant Analysis. (A) Group averaged performance of a linear classifier discriminating between stimulus type (auditory: high vs. low tone; visual: small vs. large circle). Auditory classification performance shown in blue, visual classification shown in orange. Horizontal dotted lines represent the threshold for statistical significance (FWER $p<0.01$, blue for auditory, orange for visual). Vertical dotted line represents the last time point (545 ms) both auditory and visual discrimination are significant (used as a time window cut-off for following analysis). Scalp topographies display the forward models (correlation between discriminant output Y and underlying EEG activity) for a 30ms time window centred on each peak in classification performance (auditory first peak: 105ms, second peak: 180 ms; visual first peak: 170 ms, second peak: 235 ms). (B) Group averaged discriminant output (Y) for auditory trials separated by congruency. Black dots represent significant congruency effect ($p<0.05$, cluster randomisation test). Scalp topographies represent the group-averaged raw EEG activity underlying this significant time window of interest (averaged across time) for congruent and incongruent trials separately. Bottom scalp topography (group averaged, time window averaged, raw EEG data) shows the difference in raw EEG activity (Congruent – Incongruent) (C) Group averaged discriminant output (Y) for visual trials, separated by congruency, with significant time points again denoted with black circles. All scalp topographies as in (B), but for the first cluster showing significant differences in the visual modality. Note: we have only included the topographies underlying the first cluster here as this was the only one which occurred during stimulus presentation. To see scalp topographies for the final two significant clusters (cluster 2,3 occurring after stimulus offset), see Supplementary Figure 1. (D) Group-level regression weights (beta values) for the regression of single trial RTs against the discriminant output for auditory (left) and visual (right). Significant time points represented with black circles ($p<0.05$, cluster permutation test). In all figure panels, grey background represents the time period of stimulus presentation (0 ms to 300 ms).

correlations between the discriminant output (Y) and the EEG activity over posterior and frontal regions (for the first classification peak and second classification peak respectively). These results indicate that the linear classifier could identify EEG components carrying significant task-relevant sensory information, and that these components were different for auditory and visual trials.

Figure 5.3B (left) shows the discriminant output (Y) for the auditory modality, separately for congruent and incongruent trials (with the signal Y flipped in order to map all trials to positive values, *see* Methods). For the auditory modality we found one cluster where the discriminant output (Y) was significantly different for congruent compared to incongruent trials (cluster randomisation test, cluster 1: $t = 8.576$, $p<0.001$, effect size = 0.4599). This emerged early in the trial, from 95 ms to 110 ms after stimulus onset, and revealed there was more information (i.e. higher value decoding signal) about stimulus type in the congruent signals compared to incongruent signals. The corresponding scalp topographies represent the raw EEG amplitude underlying this time window of interest, shown separately for congruent (Figure 5.3B, top row, left) and incongruent trials (Figure 5.3B, top row, right). These revealed

similar activation patterns over posterior-temporal regions for auditory congruent and incongruent trials. However, examining the difference (congruent-incongruent) between these raw EEG topographies (Figure 5.3B bottom row) revealed positive differences over frontal, temporal and posterior regions.

Figure 5.3C (left) displays the discriminant output (Y) for the visual modality, separately for congruent and incongruent trials (again, with the Y signal flipped to map all trials to positive values, *see* Methods). There were three clusters where the discriminant output was modulated by congruency. The first emerged from 220ms to 235ms (cluster randomisation test, cluster 1: $t(18) = -7.761$, $p<0.001$, effect size = 0.426). The corresponding scalp topographies displaying the raw EEG activity underlying this time window of interest revealed strong occipital activity for both congruent and incongruent trials (Figure 5.3C top row, left and right respectively). Examining the difference between congruent and incongruent trials revealed strong positive differences over frontal and temporal electrodes. The second cluster emerged after stimulus offset at 460ms to 465ms (cluster randomisation test, cluster 2: $t(18) = 4.0906$, $p<0.001$, effect size = 0.441). At this cluster, there was strong negative activity over frontal electrodes for both congruent and incongruent trials (Supplementary Figure 5.1A, top row, left and right respectively). Examining the difference (congruent – incongruent) between these topographies showed a central effect (Supplementary Figure 5.1A, bottom row). Finally, the third significant cluster which showed differences based on visual congruency occurred from 530ms to 545ms (cluster randomisation test, cluster 3: $t(18) = 9.038$, $p<0.001$, effect size = 0.480). This time, the cluster was associated with strong negative activity over frontal and temporal electrodes (Supplementary Figure 5.1B, top row, left and right respectively). Examining the EEG signal difference between congruency conditions again revealed–a weaker positive difference over frontal regions between congruent and incongruent trials (Supplementary Figure 5.1B, bottom row).

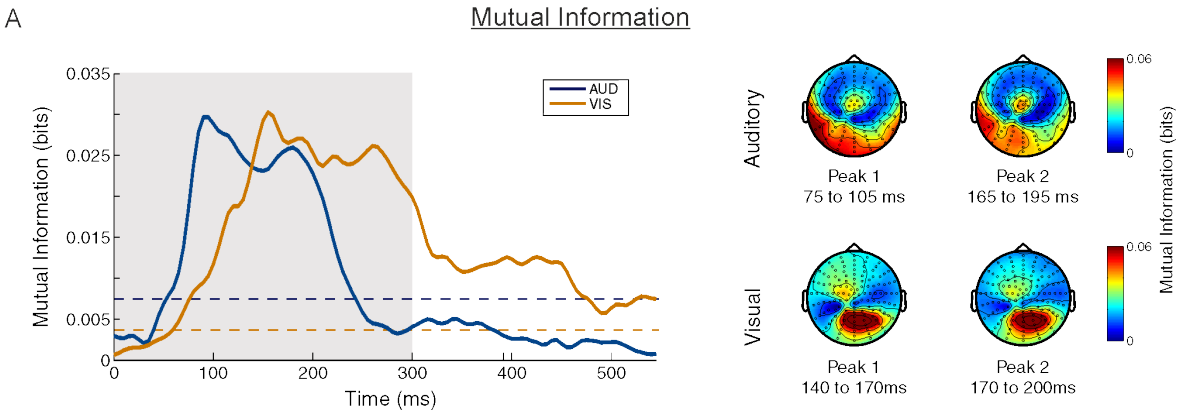
Figure 5.3D (left) displays the results generated from regressing the auditory decoding signal against auditory trial reaction times. Here we found that the discriminant EEG signals were significantly predictive of reaction times at two clusters during stimulus presentation: from 65 ms to 135 ms, and from 180 ms to

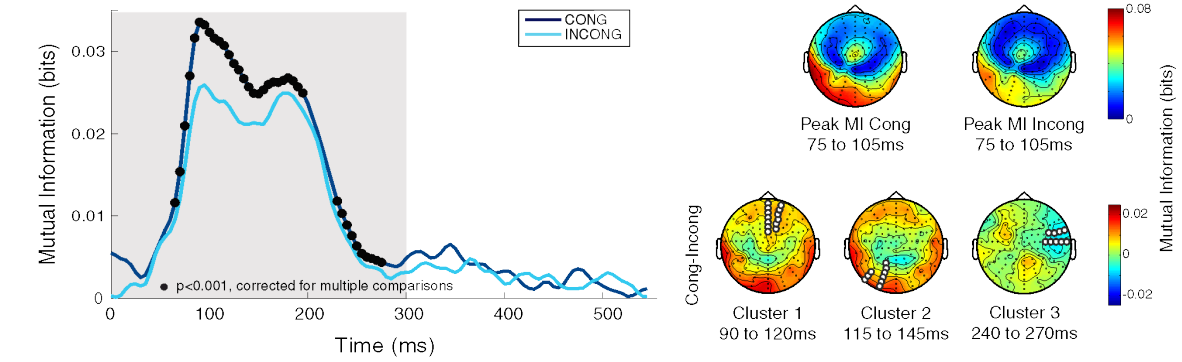
205 ms (cluster randomisation tests: first cluster, $t(18) = 40.479$, $p<0.001$, effect size = 0.541; second cluster, $t(18) = 14.538$, $p<0.001$, effect size = 0.504). Figure 5.3D (right) displays result from regressing the visual decoding signal against visual trial reaction times. Here, we found that visual discriminant signals were not predictive of reaction times at any point in the trial (cluster randomisation test, $p>0.01$).

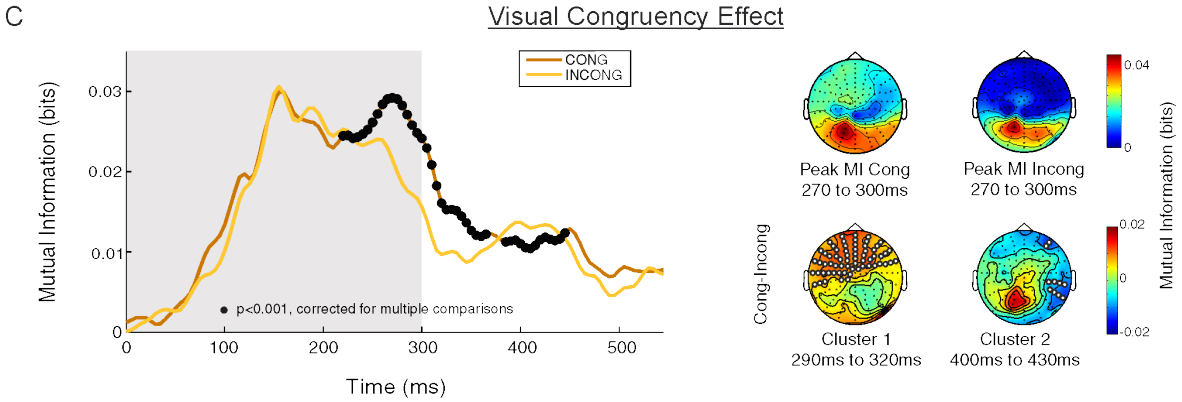
To summarise the LDA analysis, the results indicated possible differences in the contributions of EEG activity to sensory discrimination and congruency due to the varying time windows and locations of effects. They also demonstrated that congruency effects emerge early during the auditory trials (during stimulus presentation) yet later during visual trials (around stimulus offset and closer to the behavioural response). Finally, the results showed that auditory discriminant EEG signals are predictive of behavioural response times, whilst visual signals are not.

Mutual Information

Figure 5.4A shows the results from the mutual information analysis, for auditory and visual trials separately. For auditory trials, stimulus information (high/low tone) was represented in the EEG signal early in the trial, with significant MI values emerging from 50 ms to 245 ms after stimulus onset for the congruent trials, and from 70 ms to 230 ms for the incongruent trials (Figure 5.4A, blue horizontal dotted line, randomisation test, 99^{th} percentile). This information was highest over posterior and temporal electrodes (Figure 5.4A, topographical inserts, top row). For the visual trials, stimulus information (small/large circle) was represented in the EEG signal early in the trial, with significant MI values emerging from 70ms to 545ms for congruent trials and from 60ms to 545ms for incongruent trials (Figure 5.4A, yellow horizontal dotted line, randomisation test, 99^{th} percentile). This time, the highest information was centred over posterior electrodes (Figure 5.4A, topographical inserts, bottom row). Note that as MI is an unsigned quantity the signs of the values are different, but the spatial patterns obtained from the sensor-wise MI analysis are very similar to the absolute value of the patterns obtained through the LDA forward model (Figure 5.3A).


A
Mutual Information
Mutual Information (bits)
0.035
0.025
0.015
0.005
0
AUD
VIS
0
100
200
300
400
500
Time (ms)
Auditory
Visual
Peak 1
75 to 105 ms
Peak 2
165 to 195 ms
Peak 1
140 to 170ms
Peak 2
170 to 200ms
0.06
0
Mutual Information (bits)


B
Auditory Congruency Effect
Mutual Information (bits)
0.03
0.02
0.01
0
CONG
INCONG
p<0.001, corrected for multiple comparisons
0
100
200
300
400
500
Time (ms)
Peak MI Cong
75 to 105ms
Peak MI Incong
75 to 105ms
Cong-Incong
Cluster 1
90 to 120ms
Cluster 2
115 to 145ms
Cluster 3
240 to 270ms
0.08
0
0.02
-0.02
Mutual Information (bits)


C
Visual Congruency Effect
Mutual Information (bits)
0.03
0.02
0.01
0
CONG
INCONG
p<0.001, corrected for multiple comparisons
0
100
200
300
400
500
Time (ms)
Peak MI Cong
270 to 300ms
Peak MI Incong
270 to 300ms
Cong-Incong
Cluster 1
290ms to 320ms
Cluster 2
400ms to 430ms
0.04
0
0.02
-0.02
Mutual Information (bits)

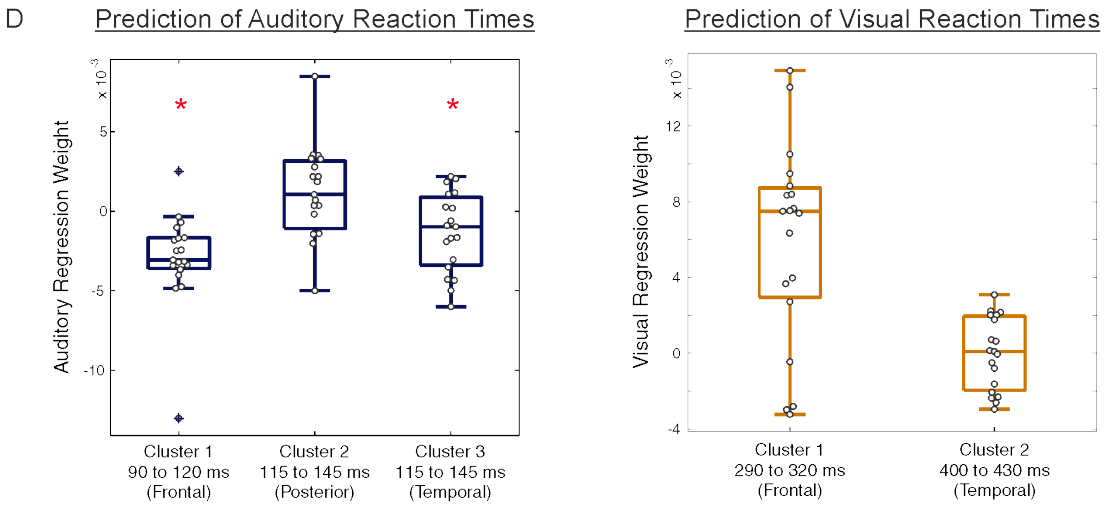

D
Prediction of Auditory Reaction Times
Auditory Regression Weight
x 10^-3
5
0
-5
-10
Cluster 1
90 to 120 ms
(Frontal)
Cluster 2
115 to 145 ms
(Posterior)
Cluster 3
115 to 145 ms
(Temporal)
Prediction of Visual Reaction Times
Visual Regression Weight
x 10^-3
12
8
4
0
-4
Cluster 1
290 to 320 ms
(Frontal)
Cluster 2
400 to 430 ms
(Temporal)

Figure 5.4 | EEG Mutual Information Analysis. (A) Group level mutual information (MI) between EEG and stimulus type (auditory: high vs. low tone, visual: small vs. large circle) averaged over electrodes. Auditory MI shown in blue, visual MI shown in orange. Dotted lines show time windows where the MI for each condition was significant (blue dotted line = auditory significance, orange dotted line = visual significance). Scalp topographies (right) show the MI at peak time points for auditory (top row) and visual (bottom row) separately. (B) Congruency difference between congruent and incongruent auditory MI (averaged over trials and electrodes). Significant time points where there was a difference between congruent and incongruent MI represented with black circles ($p<0.05$, cluster randomisation test). Scalp topography (B, top row) shows MI underlying congruent (left) and incongruent (right) auditory trials underlying peak MI (averaged over 30ms around centre of peak MI difference). Scalp topography (B, bottom row) shows the MI difference (Congruent – Incongruent) for the three clusters where there was a significant congruency effect (again, topographies are activity averaged over 30ms around centre of each significant cluster). (C) Same as in (B), but for visual trials. (D) Regression weights (beta values) generated from regressing EEG activity against reaction time for the three clusters of interest for auditory (left) and visual (right) trials. Grey circles represent individual subject beta weights, and asterisks represent clusters where EEG activity significantly predicted reaction time. Again, in all panels grey background represents the time of stimulus presentation (0 ms to 300 ms).

Figure 5.4B (left) displays the results from the congruency comparison for auditory trials, with congruent and incongruent trials shown separately. This revealed stronger MI between the EEG signal and stimulus for congruent compared to incongruent trials, emerging early in the trial. The scalp topographies show that the highest MI for both congruent and incongruent trials occurred over posterior and temporal electrodes (Figure 5.4A, topographical inserts, top row, MI averaged over a 30 ms time window centred on peak). Comparing MI values between congruent and incongruent auditory trials revealed three significant spatio-temporal clusters: one emerging from 65 ms to 145 ms over frontal electrodes (cluster randomisation test, cluster 1: $t(18) = 160.1086$, $p = 0.004$, effect size = 0.4176), one emerging 65 ms to 195 ms over posterior electrodes (cluster randomisation test, cluster 2: $t(18) = 180.649$, $p = 0.004$, effect size = 0.4389), and one from 230 ms to 275 ms over temporal electrodes (cluster randomisation test, cluster 3: $t(18) = -131.843$, $p < 0.001$, effect = 0.4751). To avoid artifactual effects in the scalp topographies occurring a result of averaging over these long significant cluster time windows, we reduced the window length of the significant clusters to a 30 ms time window centred on each MI cluster (cluster 1 centre, 105ms; cluster 2 centre, 130ms; cluster 3 centre, 250ms). Figure 5.4B (topographical inserts, bottom row) displays the early differences found over frontal (left), temporal (middle), and posterior (right) electrodes for each shorter spatio-temporal cluster (Figure 5.4B, topographical inserts, bottom row).

Figure 5.4C (left) displays the results from the congruency comparison for visual trials, with congruent and incongruent trials shown separately. Again, this revealed stronger MI between the EEG signal and stimulus for congruent compared to incongruent visual trials. The scalp topographies indicated that MI between the EEG signal and stimulus type was strongest over posterior electrodes (Figure 5.4C, topographical inserts, top row, MI averaged over a 30ms time window centred on peak). In contrast to the auditory modality, comparing congruent and incongruent visual MI revealed significant differences at two spatio-temporal clusters later in the trial: one from 220 to 365ms (cluster randomisation test, cluster 1: $t(18) = 250.67$, $p < 0.001$, effect = 0.4848) over frontal and temporal electrodes, and one from 385 to 445ms (cluster randomisation test, cluster 2: $t(18) = -99.309$, $p < 0.001$, effect = 0.4309) over temporal electrodes. As for the auditory, the time window for displaying each spatio-temporal cluster was again reduced to a 30 ms time window centred on each cluster which showed significant differences based on congruency (cluster 1 centre, 290ms; cluster 2 centre, 415ms). Figure 5.4C (topographical inserts, bottom row) displays the differences found over frontal (left) and temporal (right) electrodes for each shorter spatio-temporal cluster.

Finally, Figure 5.4D displays the results of the regression analysis between the EEG activity underlying the significant clusters and single-trial reaction times for auditory (left) and visual (right) separately. As a reminder, to calculate the regression weights we averaged the EEG activity over all significant electrodes (shown in Figure 5.4B and 5.4C, white circles) and over the shorter, 30ms windows centred over the cluster peaks (defined above) and regressed the activity in each cluster against behavioural reaction times. For auditory trials, this regression analysis demonstrated that the EEG in both the early frontal and later temporal cluster was predictive of reaction times (cluster 1: 90 ms to 120 ms, $t(18) = -4.3073$, $p = 0.00042$, effect size = -0.988; cluster 3: 235 ms to 265 ms, $t(18) = -2.2979$, $p = 0.0338$, effect size = -0.5272 for frontal and temporal respectively). However, auditory activity in the posterior cluster was not significantly predictive of reaction times (cluster 2: 115 ms to 145 ms, $t(18)$, $= 1.349$, $p = 0.194$, $d = 0.309$). For the visual trials, the EEG activity in both the frontal and temporal clusters was not significantly predictive of visual trial reaction times.

To summarise, the MI analysis results indicated that the encoding of acoustic information was affected by congruency early in the trial, and activity over frontal and temporal electrodes was significantly predictive of reaction times at these early latencies. In contrast, the encoding of visual information was affected by congruency only later in the trial over frontal and temporal electrodes, and this activity was not significantly predictive of reaction times.

Consistency of Results

Both EEG analyses revealed that the EEG signal contained information about stimulus type during stimulus presentation, with the highest information emerging over posterior and temporal regions for auditory trials, and over posterior regions for visual trials (see Figure 5.3A and Figure 5.4A). In both analyses, effects of stimulus congruency emerged early in the trial for auditory (Figure 5.3B and Figure 5.4B), and the strongest differences appeared over posterior, temporal and frontal regions. In both analyses, effects of stimulus congruency for visual trials emerged later in the trial with the strongest differences over frontal and temporal regions (Figure 5.3C and Figure 5.4C). Finally, both analyses demonstrated that neural activity was predictive of behavioural reaction time during stimulus presentation for auditory trials, and that neural activity underlying visual trials was not predictive of behavioural reaction time (see Figure 5.3D, Figure 5.4D). Overall, these similar findings across methods are encouraging, and demonstrate that our results are robust.

Discussion

In this experiment we examined the neural mechanisms underlying a cross-modal association between auditory pitch and visual size. To do so, we used an implicit association test (IAT) in combination with 128-channel EEG recording and two approaches to multivariate EEG analysis. In contrast to previous work using the IAT (Parise & Spence, 2012), we manipulated only the assignment of the auditory stimuli to response keys and held the visual stimulus-response key mapping constant. Consequently, we hypothesised that we would see an effect of reaction time based on perceived congruency for auditory trials only. As predicted, the behavioural data revealed that subjects responded faster on auditory congruent trials than incongruent trials, providing evidence for an

implicit association. From the EEG data we found that information about the stimulus could be extracted from the EEG activity during stimulus presentation for both auditory and visual trials, and that effects related to stimulus type were strongest over posterior and temporal regions for auditory, and posterior regions for visual. The results also revealed that auditory neural correlates were modulated by congruency early in the trial (<100 ms), and that these differences emerged from frontal, posterior and temporal regions. Visual neural correlates were modulated by congruency later in the trial (>250ms) over frontal and temporal areas. Finally, auditory activity underlying a congruency modulated representation of the auditory stimulus in frontal and temporal regions was also significantly predictive of single trial auditory reaction times, indicating that these activations reflect a neural correlate of the underlying perceptual association. In contrast, activity underlying a congruency modulated visual stimulus representation was not significantly predictive of visual reaction times. Importantly, these results were consistent across two separate analysis methods, showing our results are robust. As a result, our data provide support for cross-modal associations as an early perceptual process that emerges at a sensory stage of perception, rather than a process having an origin during a later decision related stage.

Effect of Cross-modal Congruency on Behaviour

The behavioural results showed that subjects had faster reaction times for auditory congruent stimulus-response assignments than incongruent ones. This provides further evidence supporting the existence of an acoustic pitch – visual size cross-modal association, which has been reported in various experimental paradigms before (Bien, ten Oever, Goebel, & Sack, 2012; Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, Ben-Artzi, & Lakatos, 2003; Parise & Spence, 2008; Parise & Spence, 2009). More importantly, given our choice of the IAT, our behavioural results demonstrate that a cross-modal association between pitch and size arises can occur even when only a single stimulus is presented on a trial. This finding replicates the work using the IAT carried out by Parise & Spence (2012). Furthermore, as the IAT presents only one stimulus per a trial, it rules out the possibility that our effects are due to general multisensory benefits (i.e. due to spatial and temporal congruency), or to attentional differences caused by dividing attention between two stimuli.

Early Neural Correlates of Pitch-Size Association

Our results revealed that auditory neural signals were modulated by the perceived congruency of sensory information early during the trial (<100 ms after stimulus onset), while visual trials were modulated later (>220 ms). The early onset of these results suggests that these implicit cross-modal associations arise at an early, perceptual stage rather than a later decision related stage. Supporting this interpretation, a recent study by Mostert, Kok, & de Lange, (2015) showed that decision related neural correlates can be dissociated from sensory related correlates by examining the temporal profile of each. Using MEG, two tasks (one sensory, one sensory and decision making), and a dual decoding approach, they demonstrated that sensory related processes emerged in occipital areas from 130 ms, whereas decisional related processes emerged later, around 250 ms. These timings are consistent with our work, with both auditory and visual modulations occurring earlier than 220ms after stimulus onset.

Additionally, the timing of our results are in line with previous neuroimaging studies examining cross-modal effects in the brain. For example Kovic et al., (2010) found neural signals were modulated around 140 ms to 180 ms by congruency (sound-symbolic association), while Bien et al., (2011) found effects emerging at 250 ms. The short latency onset of these effects occurring in response to the presentation of two associated arbitrary stimulus properties led both sets of authors to conclude that these associations arise in the early stages of the multisensory integration process. These claims of cross-modal associations originating due to multisensory integration are in line with past work that has shown multisensory interactions occurring in neural signals at very short latencies after stimulus onset (Giard & Peronnet, 1999; Molholm et al., 2002; Sperdin et al., 2009). However, in our study the early modulations we observe cannot be due to multisensory integration, as on each trial only a single stimulus was presented. Instead, the modulations we see – while multisensory in nature – must rely on a different perceptual process. We suggest that the congruency effects we observed may be due to some form of top-down feedback influencing signals in the different modalities at an early stage during the perceptual process, or from some existing underlying mapping of a pitch-size association which automatically influences early sensory processing. In further support of

this, the early onset of our effects emerged during stimulus presentation, and at least 600 ms before any decisional response is made. This timing difference would appear to rule out explanations of the results in terms of decision related strategies. This evidence further strengthens our interpretation effects of a cross-modal association arise early during sensory processing rather than at a later decisional stage.

We also observed that neural correlates of auditory congruency emerge early (<100ms) and are predictive of behavioural reaction times, while visual correlates emerged later (>200ms) and are not predictive of behavioural reaction times. This dissociation is interesting, given that we see an effect of congruency on behaviour for auditory trials, but not for visual trials. Together, this potentially indicates that the auditory modulations are more behaviourally relevant than the visual. A first potential explanation for this dissociation in both temporal profiles and predictive validity of the signals is that the auditory stimuli changes on a trial-to-trial basis, the encoding of congruency may have to be encoded early than the visual trials (which do not change). A second potential interpretation is that the auditory and visual neural correlates possibly relate to different underlying processes, which emerge with different temporal profiles (as in the Mostert et al., (2015) study mentioned above). Specifically, while neural signals are modulated by congruency regardless of modality, these modulations may only be behaviourally relevant and affect subsequent perceptual decisions if they occur early after stimulus onset. With regards to our results, the early auditory modulations could represent the updating of audio-visual congruency (as the congruency of the auditory stimuli changed on a trial-to-trial basis), whereas the visual modulations could represent a stable decisional related correlate based on visual congruency. The timing of the auditory (<100ms) and visual (>200ms) would also fit with the Mostert et al., (2015) study showing a dissociation between early perceptual related correlates (<100ms) and later decision related correlates (>200ms).

A final interpretation of the early effect could relate to the nature of the association itself being environmentally relevant. For example, an acoustic pitch - visual size association reflects the natural properties of acoustic resonance in the world where larger objects resonate at lower frequencies than smaller ones. Other cross-modal associations, such as acoustic pitch – spatial

elevation, also rely on a strong environmental relationship: larger objects tend to be heavier and more likely to be found at lower elevations than smaller, lighter objects. Indeed, a recent study has even shown that both auditory scene statistics in sounds recorded from the natural world, and the filtering properties of the human ear also demonstrate a clear mapping between frequency and elevation; specifically, high-frequency sounds tended to originate from high elevation sources in natural auditory scenes, and sounds coming from low elevations had less energy than those coming from high elevations (Parise, Knorre, & Ernst, 2014a). Given that such cross-modal associations reflect a naturally occurring link between stimuli, this could lead to a strong Bayesian prior on this relationship (de-Wit, Machilsen, & Putzeys, 2010; Huang & Rao, 2011; Knill & Pouget, 2004). Future work could attempt to investigate such relationships between implicit cross-modal associations and prior probabilities of co-occurrence through experiments that manipulate the priors, or involve more arbitrary pairings of stimuli (e.g. colour and pitch).

Spatial Distribution of Early Cross-Modal Effects

Both the LDA and MI analysis revealed that congruency differences emerged over frontal, temporal and posterior areas for auditory trials, and frontal and temporal electrodes for visual trials. This points towards a dynamic network of areas underlying the implementation of these implicit associations in order to affect subsequent behavioural responses. These findings are also consistent with other studies investigating how neural activity is modulated during cross-modal associations. For example, Bien et al., (2012) found modulations in ERPs over frontal regions, and using fMRI Sadaghiani, Maier, & Noppeney, (2009) demonstrated that higher-level speech-motion cross-modal associations emerged in fronto-parietal areas.

However, modulations in parietal activity are the most common finding across all previous studies examining the neural underpinnings of cross-modal associations. For example, Bien et al., (2012) found that parietal ERPs (200 ms - 300 ms) were modulated by congruency, and that TMS applied over parietal cortex reduced the amplitude difference between congruent and incongruent trials, and the behavioural congruency effect. Similarly, Kovic,

Plunkett, & Westermann, (2010) found early ERP modulations over parietal sites, and — as mentioned above — Sadaghiani, Maier, & Noppeney, (2009) found an interaction in fronto-parietal areas in response to audio-visual motion stimuli. In our study we found no parietal activity which was predictive of behaviour or modulated by congruency, and so our findings are at odds with past results in the literature.

However, the findings of these previous studies are confounded with the issue of multisensory integration and attention; in all experiments, two stimuli were presented on each trial (either simultaneously or sequentially). Previous studies have found early parietal effects related to bimodal stimuli presentation (Molholm et al., 2002; Murray et al., 2005), and the parietal region is often implicated in multisensory integration (Matsuhashi et al., 2004; Molholm, 2006; Pasalar et al., 2010; Rohe & Noppeney, 2015a, 2016). Additionally, parietal activity has been shown to be involved in attention during multisensory tasks (Downar, Crawley, Mikulis, & Davis, 2001, 2002; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). Therefore, parietal activity observed in these previous studies may be reflecting multisensory processing or attention, rather than reflecting a pure cross-modal association. As a result, the effects of parietal TMS on cross-modal effects in the Bien et al., (2012) study might simply disrupt multisensory integration processes rather than specific cross-modal association processes, and the parietal activation seen in the Kovic et al., (2010) and Sadaghiani, Maier & Noppeney (2009) study may reflect audio-visual processing or divided attention. In contrast, in our experiment we presented only a single stimulus on each trial, thus ruling out effects of multisensory integration or divided attention. We propose this is stronger evidence that the parietal component represents effects due to multisensory integration or attention in neural signals rather than the cross-modal association, and that the effects we find here may be a more robust correlate of cross-modal associations in the brain.

This interpretation is strengthened by our finding that auditory neural signals in frontal and temporal regions were significantly predictive of reaction time early during the trial (i.e. during stimulus presentation) across two analyses methods. Thus, we directly show that specific EEG components in frontal and temporal areas are behaviourally relevant, with the addition of showing this is robust

across two separate analysis methods. This supports our interpretation that the effects seen in our study over temporal and frontal regions may be a more sensitive measure of where in the brain a cross-modal association first emerges, and that parietal activity found in previous studies may be more reflective of other perceptual and cognitive processes. Additionally, as the neural signals are behaviourally relevant early in the trial, the results once again support the early and perceptual account of cross-modal associations rather than the later, decisional account.

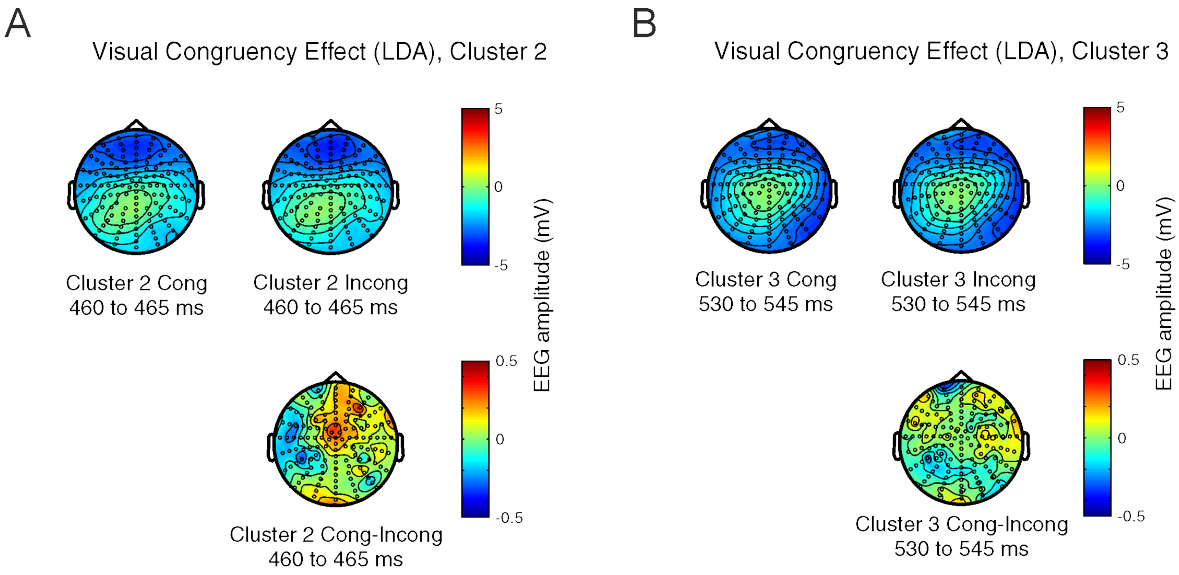
Consistency of Results

A final aim of this work was related to the analyses methods. In recent years, it has been suggested that the wide range of different processing pipelines and alternative analyses methods leads to heterogeneity in neuroimaging results (Carp, 2012a, 2012b). This can lead to divergent findings and so difficulty interpreting the overall literature on a specific topic. To address this and reveal how different analyses pipelines would affect the results from the EEG data we collected, we used two EEG analysis methods (linear discriminant analysis and mutual information analysis) to analyse the same experimental data set. This allowed us to examine whether our results were robust by looking for consistency across methods. We found that across analysis methods, the timing and locations of the results were consistent, suggesting our results are robust to the choice of analysis method.

Chapter 5: Summary

This chapter has presented results showing that multisensory interactions (in the form of a cross-modal auditory pitch-visual size association) emerge early during the perceptual process. Once again, this effect was demonstrated using a combination of neuroimaging, psychophysics, and single-trial analysis methods, again stepping beyond examining simple modulations in response to stimuli presented. Finally, this chapter — although examining multisensory interactions instead of multisensory integration — aligns with the previous two chapters by providing the third set of results supporting the early and distributed view of multisensory processing.

Supplementary Materials: Chapter 5



Supplementary Figure 5.1 | Significant visual LDA clusters (after stimulus offset). (A) Scalp topographies represent the group-averaged raw EEG activity underlying the second significant cluster (cluster 2) where congruency differences were found, averaged across time for congruent and incongruent trials separately. Bottom scalp topography (group averaged, time window averaged, raw EEG data) shows the difference in raw EEG activity (Congruent – Incongruent). (B) Shows the same as (A) for the third significant visual cluster.

Chapter 6 : General Discussion

The work in this thesis investigated the temporal dynamics underlying audio-visual perception in the brain using a combination of psychophysics, high-density neuroimaging (EEG), and multivariate analysis methods. This novel combination made it possible to extract information from the EEG signal and relate it to behaviour on a trial-by-trial and time resolved basis across three experiments. As a result, the work presented here steps beyond previous studies to provide new insights into the temporal dynamics of audio-visual perception in the brain during a variety of perceptual tasks.

Chapter 3 (Experiment 1) focused on investigating when neural correlates of sensory reliability and perceptual weighting emerged during audio-visual integration. The behavioural results showed that subjects weighted sensory information in a manner broadly consistent with the rules of optimal integration: as visual reliability decreased, auditory perceptual weights increased. However, they also revealed a consistent bias towards the auditory modality regardless of visual reliability, which caused a mismatch between the perceptual weights predicted by the model and the perceptual weights observed in subject behaviour. We exploited this dissociation between sensory reliability and perceptual weighting to uncover neural correlates related to each, using a combination of single trial analysis and regression modelling. These results demonstrated that information about sensory rate could be extracted from neural signals during stimulus presentation, with the strongest information over scalp locations consistent with early sensory regions. This information was modulated by visual reliability early during the perceptual process in sensory and parietal regions. Similarly, neural correlates of perceptual weighting emerged in neural signals early after stimulus onset, again in early sensory and parietal regions. Overall these results broadly support previous findings from the fMRI literature which have indicated cue weighting occurs across the cortex (Beauchamp et al., 2010; Helbig et al., 2012; Rohe & Noppeney, 2016), and provide the first demonstration of the temporal dynamics of sensory cue weighting in the human brain.

Chapter 4 (Experiment 2) expanded on the results from Chapter 3. Specifically, this chapter investigated whether oscillatory power during audio-visual trials was modulated by sensory reliability. The results showed there was lower theta power and higher alpha power during audio-visual conditions where visual reliability was low, relative to conditions where visual reliability was high. These modulations occurred early during the trial (0 ms to 300 ms) over fronto-central regions, and alongside significant changes in perceptual weighting. Again, these results demonstrate for the first time that theta and alpha are modulated due to reliability based cue weighting in the brain, thus adding important insights to the field.

Finally, Chapter 5 (Experiment 3) investigated when effects of an auditory pitch, visual size cross-modal association emerged in the brain. The behavioural results replicated findings from previous work showing that participants responded faster when congruent pairings of auditory and visual stimuli were assigned to the same response key, compared to when incongruent pairings were. The EEG data revealed that information about stimulus type could be discriminated early in the trial during stimulus presentation, with the strongest information contained in temporal and posterior areas. Neural components related to this information were modulated by the perceived congruency of the stimuli early in the trial (<100ms), over temporal, posterior and frontal areas for auditory trials, and later in the trial (>250 ms) over frontal and temporal areas for visual trials. Finally, the EEG signal over frontal and temporal regions during auditory trials was predictive of behavioural reaction times, while the activity underlying visual trials was not. Importantly these results were consistent across two analysis methods, suggesting our findings are robust. Once again, these results provide the first investigation into the neural underpinnings of an audio-visual cross-modal association, thus adding an important contribution to the field.

To sum up, the common finding across all three experiments was that early neural correlates of multisensory stimuli and behaviour were observed over sensory (posterior and temporal), association (parietal) and higher-order (frontal) regions. Given that each experimental chapter has already addressed the specific effects from each experiment, this chapter will focus on discussing the common findings across experiments. In addition, it will highlight some

conceptual issues related to the experiments and methods used, before finishing by providing suggestions for future work.

Early Effects of Multisensory Perception

How do our findings of early, distributed effects fit in with the field? As described in the General Introduction (Chapter 1), research has shown that multisensory interactions can occur surprisingly early during perception (40-80 ms, Giard & Peronnet, 1999; Molholm et al., 2002; Murray et al., 2005) at almost every level of the cortex (for review, see Ghazanfar & Schroeder, 2006). These observations led to the modern view that multisensory perception is an early emerging process in the brain, rather a later, hierarchical one. The work presented in this thesis supports this view by demonstrating that a dynamic network of early sensory and higher order regions is active early during a variety of audio-visual perceptual tasks.

However, across experiments, our effects emerged later than those reported by previous studies. For example, enhancement of auditory ERPs in response to a simultaneously presented somatosensory stimulus have been shown to emerge as early as 50 ms after stimulus onset (Murray et al., 2005). Audio-visual ERP responses have been found to occur at between 40 ms (Giard & Peronnet, 1999) and 46 ms (Molholm et al., 2002) in occipital and parieto-occipital regions respectively. Finally, Cappe, Thut, Romei, & Murray, (2010) showed audio-visual ERP responses were modulated in primary visual, auditory and temporal areas from 60 ms after stimulus onset. In contrast, the earliest effects we found emerged at later, ranging from 65 ms to 260 ms across Chapters.

Why do we not find effects as early as previous work? One possibility is due to the questions of interest studied in these experiments. Previous studies focus on uncovering effects relating to multisensory stimuli by comparing multisensory signals to unisensory signals (Molholm et al., 2002; Murray et al., 2005). In contrast, rather than focusing on effects related to basic properties of multisensory stimuli, the analyses in this thesis examined perceptual effects related to multisensory processes, such as cue weighting, oscillatory modulation, or cross-modal interactions. For example, in Chapters 3 and 4 the aim was to

uncover neural correlates of audio-visual reliability and perceptual cue reweighting. In Chapter 5, the goal was to examine cross-modal perceptual associations between unisensory auditory and visual stimuli. It may be that comparing audio-visual trials across different experimental conditions (as we did here) does not shed light on effects related to basic stimulus processing that are evident when the analyses focuses on comparing multisensory to unisensory trials. This speculation is in part backed up by the observation that in Chapters 3 and 5, basic audio-visual stimulus features (rate, stimulus type) could be reliably decoded from stimulus onset or shortly thereafter (latest: 45 ms, Chapter 3). This indicates that information is indeed contained in the EEG signal about basic audio-visual stimulus features, and leads us to speculate that if we had compared unisensory to multisensory signals (as the previously mentioned studies do), we may have found early differences on a time-scale in line with previous work in the field.

Activity in Early Sensory Regions: Feedback or Feedforward?

Our results showed early results related to multisensory stimuli and behaviour arose in low-level sensory areas (posterior and temporal) across a variety of experiments. These results are similar to the findings of many other studies examining multisensory processing in the brain, who have found activations in the same regions in human and animal work (Foxe et al., 2000; Giard & Peronnet, 1999; Kayser & Logothetis, 2007; Kayser, Petkov, & Logothetis, 2008; Murray et al., 2005; Sperdin, Cappe, Foxe, & Murray, 2009). However despite the widespread finding of multisensory activity in sensory regions, it remains unclear what role these early areas play in multisensory processing. Low-order regions of sensory cortex may process “feedforward” information coming directly from multiple senses to early sensory areas, or they could process “feedback” (i.e. top down influences) from higher cortical areas or other sensory regions.

With regards to our results, the critical issue to consider is timing: is there sufficient time for the early effects we saw in sensory areas to have been transmitted via feedback, either from other sensory areas or higher regions of

the cortex? Research shows visual processing in occipital cortex occurs around 50 ms after stimulus onset (Clark, Fan, & Hillyard, 1994; Schroeder, Mehta, & Givre, 1998), and auditory processing in auditory cortex occurs around 15 ms after (Rupp et al., 2000). Taking into account what is known about the speed of neural transmission, Foxe & Schroeder, (2005) approximated the timing of feedback from higher areas to early sensory areas to occur 67 ms after stimulus onset. Under this estimate, effects due to feedforward processing should occur prior to 67 ms, while effects due to feedback processing should occur after.

With this in mind, we propose that our earliest effects in sensory regions are due to feedforward mechanisms. In Chapter 3, information about stimulus rate was decodable from the EEG signal from 48 ms after stimulus onset, with the strongest activity emerging over scalp regions consistent with sensory areas. In Chapter 5, information about acoustic pitch and visual size was decodable from 50 ms after stimulus onset, again from regions suggesting origin in early sensory cortices. Overall, the immediate onset of these effects strongly suggests that low-level encoding of basic stimulus features (such as stimulation rate, pitch and size) happens via direct input mechanisms to sensory regions.

In contrast, we speculate that later effects in sensory areas are due to feedback mechanisms. For example, in Chapter 3, effects of reliability emerged at 84 ms and 252 ms in temporal and posterior areas. In Chapter 5 effects of congruency emerged between 65 and 95 ms over temporal and posterior areas. Overall, these timings are consistent with the feedback timings proposed by Foxe & Schroeder (2005) (i.e. >67 ms). This provides a strong indication that feedback mechanisms underlie these early effects in sensory cortices. Secondly, although sensory reliability could be considered a “low level” stimulus feature, it may need to be encoded via feedback; after all, to integrate sensory cues based on reliability it is necessary to know the reliability of both cues in order to weight each appropriately. Thus, it could be that feedback mechanisms send information about sensory reliability between sensory cortices or via higher order regions early in the trial to allow accurate cue weighting. Additionally, in Chapter 5 we did not present two stimuli at the same time, and therefore the effects of one modality on the other cannot arise due to direct input to the respective cortices, but rather must emerge due either to feedback from association areas (i.e. parietal cortex), or from cross-talk between sensory

regions. Taking these interpretations together, we propose that across experiments, basic stimulus feature encoding occurs from onset due to feedforward inputs, while reliability and congruency effects in early areas occur via feedback.

Activity in Association and Higher Order Regions

Across experiments we found activity emerging in parietal and frontal regions. Given that parietal and frontal activations have previously been linked to perceptual, attention and decision-related processes it is worth examining how the effects we found relate to each of these potential interpretations.

Perception

In Chapter 3 neural correlates of sensory reliability and perceptual weighting emerged in parietal cortex. In Chapter 4, oscillatory theta and alpha power was modulated over frontal regions at the same epoch where there was a significant change in perceptual weighting. Finally, in Chapter 5, neural signals over frontal regions were modulated by audio-visual congruency, and were predictive of behavioural reaction times. Consistent with our findings of effects in higher order and association areas, many studies examining multisensory perception have found parietal and frontal activation. For example, intracranial EEG studies have found neurons in the superior parietal lobule (SPL) (Molholm, 2006) and temporo-parietal junction (Matsuhashi et al., 2004) display increased firing in response to multisensory stimuli compared to unisensory stimuli. Neuroimaging studies have shown parietal and frontal activation in response to audio-visual speech (Calvert, 2001) and basic audio-visual stimuli (Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). Finally, modulations in oscillatory activity in response to audio-visual stimuli have been found in parietal and frontal areas (Sakowitz et al., 2000; Sakowitz et al., 2005). Taken together with the literature, our results of parietal and frontal activation across experiments using audio-visual stimuli strongly suggest these areas play a general role in multisensory perception. Additionally, across experiments we demonstrated that frontal and parietal activity patterns were related to behaviour, early during perceptual tasks. These direct associations with behaviour strengthen our

interpretation that the activations seen in parietal and frontal regions in our experiments are linked to perceptual processes.

Attention

Parietal and frontal activations have also been widely linked to attention. For example, an fMRI study by Downar, Crawley, Mikulis, & Davis, (2001) showed the BOLD response in the temporoparietal junction (TPJ) was enhanced when behaviourally relevant auditory and visual stimuli were presented, as compared to when behaviourally irrelevant stimuli were. Using EEG, Shomstein, Kravitz, & Behrmann, (2012) demonstrated that cued shifts of attention resulted in modulations over frontal and parietal regions early in the trial (146 ms and 286 ms respectively). Markett et al., (2014) showed similar fronto-parietal activation during attentional tasks using fMRI. Finally, Talsma & Woldorff, (2005) demonstrated that ERP activity was stronger during attended audio-visual trials compared to unattended trials, and that these differences emerged early in the trial (100 ms, 160 ms, and 300 ms) over fronto-central regions. In Chapter 3 we found parietal activity was related to sensory reliability and perceptual weighting from 156 ms. In Chapter 4, modulations in relative theta and alpha power occurred over frontal regions from stimulus onset to 300 ms. Finally, in Chapter 5, frontal activation was modulated by audio-visual congruency from 90 ms to 120 ms. The location and timing of these effects somewhat correspond to the results in this thesis. For example, In Chapter 3 we found parietal activity was related to sensory reliability and perceptual weighting from 156 ms. In Chapter 4, modulations in relative theta and alpha power occurred over frontal regions from stimulus onset to 300 ms. Finally, in Chapter 5, frontal activation was modulated by audio-visual congruency from 90 ms to 120 ms. Given that these results from the attention literature are similar to ours, it could point towards attention underlying the effects we observed.

Unfortunately, dissociating attention processes from perceptual processes in our results is somewhat difficult. In Chapters 3 and 4, we presented two audio-visual stimuli on each trial and so the modulations we see could arise from processes recruited to allocate attention across two stimuli. However, given that the analyses focused on comparing differences between audio-visual streams — which contain equal numbers of stimuli —rather than comparing trials where two

stimuli were presented to trials where only one stimulus was, this seems unlikely. Similarly, in Chapter 5, we presented only one stimulus on each trial, and so any modulations observed in higher-order regions cannot be due to confounding attentional processes. Rather, we speculate that the frontal activation seen in Chapter 5, early in the trial represents context driven information relating to the cross-modal association, rather than attention related processes. Consequently, with regards to the work in this thesis it seems reasonable to interpret the activity seen in parietal and frontal areas as relating to perceptual rather than attentional processes.

Decision-making

Finally, studies examining perceptual decision making have also found parietal-frontal activations (Gherman & Philiastides, 2015; Paulus et al., 2001). For example, a recent study by Mostert et al., (2015) showed that subtle decision related neural correlates can be intertwined with sensory related correlates and that each has a different temporal profile. Specifically, they used MEG, two tasks (one sensory, one sensory and decision making), and a dual decoding approach, to demonstrate that sensory related processes emerged in occipital areas from 130 ms to 350 ms, whereas decisional related processes emerged later in parietal and frontal areas around 250 ms.

Whilst the location of our results corresponds to the locations found in the Mostert et al., (2015) study, the timing of our results suggests they are unlikely to be related to decision processes. In Chapters 3 and 4, parietal and frontal effects emerged from 156 ms and stimulus onset, respectively. These effects are much earlier than the decisional related effects emerged in the Mostert et al., (2015) study. Additionally, in the context of the individual experiment in Chapters 3 and 4, subjects were not required to make a perceptual decision until approximately 1.5 seconds after these parietal and frontal modulations emerged. Together, this makes it unlikely these effects were related to decision processes. Finally, in Chapter 5 frontal effects emerged from 65 ms until 145 ms, again far earlier than the divergent decision related processes reported in the Mostert et al., (2015) study. Thus — despite not performing analyses specifically designed to tease apart sensory from decisional related effects — we speculate that the results we see in higher order regions are unlikely to be related to

decision processes. To confirm, future work should focus either on using analysis aimed at separating the two processes (as in the Mostert et al., (2015) paper), or use recording techniques that can directly manipulate neural activity (such as direct neuronal recordings or brain stimulation) in order to determine causal roles of early and late effects during multisensory processing.

To summarise, we find parietal and frontal activity in two paradigms examining audio-visual perception. Taken together with the literature in the field, we speculate these effects in association and higher-order regions are related to perceptual processes. However, from our experiments it is difficult to determine conclusively whether the effects we find in association and higher regions arise from sensory, attention, or decision related processes (or a mix of all three). Future work is required to tease apart the relevant contributions.

Decoding Information from Neural Signals

Across experiments we demonstrated that information about sensory stimuli could be extracted from neural signals using multivariate methods. These results add to the growing literature using single trial analysis of EEG data to decode various stimulus features (Kayser et al., 2016; Lou, Li, Philiastides, & Sajda, 2014; Mostert, Kok, & de Lange, 2015; Philiastides et al., 2014; Philiastides & Sajda, 2006; Ratcliff et al., 2009; Wyart, de Gardelle, Scholl, & Summerfield, 2012). Additionally, our results add new insights to this field by revealing the dynamic correlates of two previously unstudied audio-visual combinations (sensory rate, cross-modal associations).

However, it is important to reflect on what we are decoding from neural signals. First, what is “information”, and what does it tell us about the brain? In the context of this thesis (and more widely, the field) “information” refers to the presence of a statistical dependence in the EEG data on the stimulus and/or experimental condition of interest. This in itself tells us a relationship of some form exists, but does it really tell us about something meaningful about the brain? For example, fluctuations in the response firing of neurons may allow a classifier to successfully discriminate between two auditory tones of different pitches. However, rather than spike rate representing the “information” the

brain uses to encode stimulus features, it may be the phase, oscillatory frequency, or spike timing that is most relevant for processing the sound and that the response firing modulations are a simple by-product of encoding. In this case, the "information" contained in the amplitude is actually meaningless, and is in fact not relevant for understanding how the brain processes auditory pitch.

This dissociation between decoding performance and neural activity was recently highlighted in a study by Kamitani & Tong, (2006). Using single-neuron recording and multivariate classification methods, they demonstrated that motion information could be decoded from early visual area V1 with more accuracy than from the higher visual area V5/MT. This was unexpected, as animal work has shown that area V5/MT primarily contains neurons which encode motion heading (Bradley, Chang, & Andersen, 1998; Krug, Cicmil, Parker, & Cumming, 2013). Overall, these findings suggest that just because we can extract information from neural signals using classifiers, it does not mean that we have found necessarily found a neural representation of our stimulus or experimental condition of interest. Taking this into account, our observation that information relating to high or low tones, small or large circles, or high and low rates could be significantly decoded from the EEG signal does not necessarily mean that we have discovered the neural representation of how or where the brain processes these stimuli. Rather, the results can only conclusively tell us that the EEG responses cluster together in a way that gives rise to successful linear separation of our conditions of interest, at early time points during the trial. Thus, rather than referring to significant clusters uncovered by decoding as "neural representations", we have endeavoured to refer to (and interpret them) throughout the text as "neural correlates".

Related to this, a failure to decode information from the brain does not mean there is no information there; it could simply be that a failure to decode (e.g. after 400 ms in Chapter 3), means only that the neural responses cluster in a non-linear way, or that the information about the stimulus contained with the signal is represented, but at a finer grain. Overall, while classifiers provide a powerful way to extract information from neural signals, we must be careful not to rely too heavily on significant classification, or discount non-significant classification as a hallmark for information processing in the brain.

One approach to provide evidence that information as uncovered by decoding is information used by the brain is to examine whether there are correlations between classifier outputs and behaviour. For example, Williams, Dang, & Kanwisher, (2007) used fMRI and multivariate pattern classification methods to examine whether the discriminant information in retinotopic and lateral occipital cortex (LOC) was higher for correct than incorrect trials. They found that while they could decode information about object category from both regions at above chance levels, only the information in the latter area was related to behaviour. Consequently, the authors argued that only the activation patterns in the LOC was actually used by the brain when forming the behavioural response (despite other regions being active).

With regards to our results, we found a similar mismatch between epochs during which classification performance was significant and those during which decoding signals were related to behaviour. For example, in Chapter 3, neural signals and behavioural responses (as measured by linear discriminant analysis) were decodable from stimulus onset to 545ms after, but only related to behaviour at two early clusters within this window. Similarly, in Chapter 5 of this thesis, information was decodable from neural signals during a large epoch (0 ms to 545 ms) over three brain areas (frontal, temporal, and posterior), but was only predictive of behavioural reaction times at two early clusters, and only over frontal and temporal regions. Together these findings strengthen our interpretations made around behaviourally relevant information extracted from neural signals, and demonstrate that quantitatively relating information (as uncovered by classifiers) to behaviour can lead to more meaningful interpretations of the brain signal and information extracted from it.

Overall, these points of discussion above do not invalidate classification as an excellent analysis technique for extracting information from neural signals; classification remains a powerful technique for studying brain function. Rather, they show us we must be careful to interpret “information” as measured by classification techniques or mutual information analysis as an important, real, and quantifiable statistical dependency, but not to overextend these interpretations to conclude that the information is a “neural representation”.

Limitations

As with all experimental work, it is important to reflect on the how the interpretations are limited by the chosen tasks and methodology. This section will reflect on three important issues to consider when interpreting the work presented in this thesis.

Modality Specificity

First, it is unclear whether the effects we find in this thesis would generalise to other modalities. For example, Chapters 3 and 4 examined audio-visual cue weighting. The results — while novel — do not tell us whether similar findings would emerge in response to auditory-tactile or visual-tactile stimuli. Speculating about this issue is complicated by the fact that to date most of the work examining cue reliability weighting in humans has focused on audio-visual stimuli (Beauchamp et al., 2010; Rohe & Noppeney, 2015a, 2016). One study by Helbig et al., (2010) examined visual-tactile cue weighting using fMRI, and found signals in primary sensory and parietal areas were modulated by reliability. This is consistent with the findings from our audio-visual study. However, since Helbig et al., (2010) used fMRI to examine visual-tactile weighting, the timing of these effects remain unclear. Thus, we can only speculate that the early neural correlates we find in our audio-visual paradigm would generalise to other modalities.

In Chapter 5 we examined an audio-visual association between acoustic pitch and visual size. While cross-modal associations have been documented to exist between stimulus attributes in many different sensory modalities — such as visual-touch (Martino & Marks, 2000), audition-touch (Yau, Olenczak, Dammann, & Bensmaia, 2009), audition-taste (Simner, Cuskley, & Kirby, 2010), and vision-smell (Gilbert, Martin, & Kemp, 1996) — to date none of these other modality combinations have been examined with neuroimaging. Therefore, again we can only speculate as to whether our results showing early modulations of neural signals would generalise to other modality combinations.

Task Generalizability

Second, would the early effects found in this thesis generalise to other tasks? In Chapters 3 and 4 we manipulated the reliability of the visual information and held auditory reliability constant. If we were to repeat the experiment and manipulate auditory reliability, we may find different results. For example, modulations in neural signals may occur earlier than we report in Chapter 3, given that auditory processing is faster than visual processing (Clark et al., 1994; Rupp et al., 2000). Similarly, in Chapters 3 and 4 the task was a temporal judgement task where the auditory modality was dominant. Would we find similar, distributed effects in a spatial localisation task where the visual modality is dominant? We speculate we might, as results from a study by Rohe & Noppeney, (2016) — which measured cue weighting using fMRI and a spatial localisation task — showed similar multisensory effects in early sensory and parietal areas. Unfortunately, as fMRI has poor temporal resolution, the timing of these effects was not clear. However, given that the effects in the Rohe & Noppeney, (2015) study emerged in areas consistent with the locations as we found in Chapters 3 and 4, we can speculate that our findings of early effects may generalise to other tasks.

In Chapter 5 we found early neural correlates using an implicit association test. Would we find similar early modulations of neural signals in a different task (e.g. one that measures direct or indirect associations?). In this case, we err on the side of caution; the cross-modal association examined here (pitch-size) may simply reflect the natural correlation between physical acoustic and visual properties of the external world whereby larger objects resonate at lower frequencies than smaller objects. Thus, this particular association may be more intrinsic, innate, and based on real world acoustic properties than other cross-modal associations are. Consequently, the effects of this particular association as measured by an implicit task may arise earlier in neural signals, than effects of a semantic association measured via an indirect tasks would.

Spatial Resolution

Finally, despite the results in this thesis providing new and important insights into the temporal dynamics underlying audio-visual perception, they do not

provide conclusive insights into the spatial localisation of these effects. As described in Chapter 2, EEG has excellent temporal resolution, but only average spatial resolution. Therefore, in this thesis inferences about the underlying locations of the effects have been made with care, evidenced by restricting our interpretations to brain regions rather than specific locations. To build on the work and make more accurate spatial interpretations alongside the temporal insights with spatial localisation, future work could use an imaging tool more suited to localisation alongside EEG. For example, combined fMRI-EEG studies have had great success in identifying the timing and location of EEG signals during perceptual (Bénar et al., 2007) and decision making (Pisauro, Fouragnan, Retzler, & Philiastides, 2017) tasks. Overall this technology potentially provides a promising way forward (if implemented carefully) to make inferences about the localisation and timing of the effects presented in this thesis.

Moving Forward

Future Experiments

The limitations listed above provide solid starting points for future work. For example, it would be interesting to investigate whether early effects in multisensory perception generalise to other modalities and tasks. Specifically, the work in Chapters 3 and 4 could be replicated using a spatial localisation task, or using visual-tactile or auditory-tactile stimuli. Similarly, the work in Chapter 5 opens many potential avenues to explore many different cross-modal associations using EEG. As mentioned in the Introduction of Chapter 5, to the best of my knowledge and research, there have only been three other studies to date investigating the neural underpinnings. Therefore there are many possible insights still to be gained.

The Importance of Linking Neural Signals to Behaviour

The primary goal of (most) cognitive neuroscience is to understand how brain signals give rise and relate to human cognition and behaviour. Consequently, interpreting neural signals without examining how they relate to behaviour (if at all) makes it difficult to evaluate whether the conclusions of the results are meaningful. The work in this thesis has shown it is possible to link neural and

behavioural signals by looking for trial-by-trial correlations between classifier outputs and behavioural signals, or by relating these signals to behaviour using regression modelling. Alternatively, work in the field has shown that modelling neural responses in the same way as behavioural responses (e.g. by fitting neurometric and psychometric curves, Fetsch et al., 2012; Gu et al., 2008) or building computational models of the brain (Knill & Pouget, 2004; Rohe & Noppeney, 2016) can provide important insights. Given the many ways it has been shown possible to carefully link behaviour and neural signals, it makes sense for neuroimaging studies to continue this process, and step away from leaning only on neural modulations or classification performance as a means of determining relevant brain activity. Overall, if one can relate neural signals to behavioural signals, any interpretations made about the data will immediately be strengthened.

Open Science

Finally, the best way to move cognitive neuroscience forward is to embrace open science. In recent years there has been a focus towards this, primarily driven forward by the failure of many studies in psychology and cognitive neuroscience to replicate (Collaboration, 2015), and problems in the field such as: low power (Button et al., 2013); erroneous and inflated significance of effects due to the dimensionality and flexibility of neuroimaging data (Gelman & Loken, 2014; Luck & Gaspelin, 2017); and "HARK-ing" – the process whereby post-hoc narratives or hypotheses are invented "after the results are known". To address these problems, open science initiatives should be widely adopted. Pre-registration and Registered Reports promote transparency of the experimental design, initial hypotheses, and planned analyses and discourage false post-hoc narratives and the temptation to carry out multiple analyses. Additionally, code sharing and open data platforms encourage accurate reporting, and allow peers to execute and evaluate the analysis and results alongside reading the publication. These efforts towards more transparent and open science — if widely accepted and encouraged — can solve the issues described above and should be promoted wherever possible. To show my support for such scientific practices, I have uploaded all data and code related to the work in this thesis to the Open Science Framework, accessible here: https://osf.io/2aexf/.

Conclusion

The work in this thesis has provided a broad investigation into the temporal dynamics underlying audio-visual perception, using advanced computational methods to extract information from the EEG signal on a single-trial basis and link these signals to behaviour in a time-resolved way. The findings of early modulations in neural correlates that link with behaviour across experimental paradigms demonstrate that audio-visual interactions can occur at all levels of the perceptual process, from early effects in sensory areas to similarly early effects in higher areas. This provides support for the prevailing modern view that the entire cortex is essentially multisensory and that multisensory effects can emerge at all stages during the perceptual process.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. http://doi.org/10.1016/j.cub.2004.01.029

Alho, K., Woods, D. L., Algazi, A., & Näätänen, R. (1992). Intermodal selective attention. II. Effects of attentional load on processing of auditory and visual stimuli in central space. *Electroencephalography and Clinical Neurophysiology*, *82*(5), 356–368. http://doi.org/10.1016/0013-4694(92)90005-3

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*(4), 452–458. http://doi.org/10.1016/j.conb.2009.06.008

Arnal, L. H., Wyart, V., & Giraud, A. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801.

Avillac, M., Hamed, S. B., & Duhamel, J. R. (2007). Multisensory Integration in the ventral intraparietal area of the macaque monkey. *Journal of Neuroscience*, *27*(8), 1922–1932. http://doi.org/10.1523/JNEUROSCI.2646-06.2007

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2000). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, *39*(2), 241–248. http://doi.org/10.1016/S0167-8760(00)00145-8

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391–1397. http://doi.org/10.1364/JOSAA.20.001391

Beauchamp, M. S. (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, *3*(2), 93–114. http://doi.org/10.1385/NI:3:2:093

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*, 809–823. http://doi.org/10.1016/S0896-6273(04)00070-4

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in Systems Neuroscience*, *4*. http://doi.org/10.3389/fnsys.2010.00025

Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and

vision in human superior temporal sulcus. *NeuroImage*, *41*(3), 1011–1020. http://doi.org/10.1016/j.neuroimage.2008.03.015

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23–23. http://doi.org/10.1167/9.5.23

Bénar, C. G., Schön, D., Grimault, S., Nazarian, B., Burle, B., Roth, M., … Anton, J. L. (2007). Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI. *Human Brain Mapping*, *28*(7), 602–613. http://doi.org/10.1002/hbm.20289

Benevento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental Neurology*, *57*(3), 849–872. http://doi.org/10.1016/0014-4886(77)90112-1

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception and Psychophysics*, *62*(2), 321–332. http://doi.org/10.3758/BF03194427

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234. http://doi.org/10.1111/j.1460-9568.2004.03670.x

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport*, *13*(14), 1727–1730. http://doi.org/10.1097/00001756-200210070-00007

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. http://doi.org/10.1016/j.neuroimage.2011.06.095

Bizley, J. K., & King, A. J. (2009). Visual influences on ferret auditory cortex. *Hearing Research*, *258*(1), 55–63. http://doi.org/10.1016/j.heares.2009.06.017

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K. R. (2011). Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage*, *56*(2), 814–825. http://doi.org/10.1016/j.neuroimage.2010.06.048

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. http://doi.org/10.1007/s00221-004-2005-z

Boylan, C., & Doig, H. R. (1988). Presaccadic spike potential to horizontal eye movements. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 559–562. http://doi.org/10.1016/0013-4694(88)90153-8

Bradley, D. C., Chang, G. C., & Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, *392*(6677), 714–7. http://doi.org/10.1038/33688

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bremmer, F., Klam, F., Duhamel, J.-R. R., Hamed, S. B., & Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *European Journal of Neuroscience*, *16*(8), 1569–1586. http://doi.org/2206 [pii]

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*(2), 369–384.

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, *10*(11), 23–23. http://doi.org/10.1167/10.11.23

Button, K. S., Ioannidis, J. P. a, Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, *14*(5), 365–376. http://doi.org/10.1038/nrn3475

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929. http://doi.org/10.1126/science.1099745

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, *11*(12), 1110–1123. http://doi.org/10.1093/cercor/11.12.1110

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., … David, A. S. (1997). Activation of Auditory Cortex During Silent Lipreading. *Science*, *276*(5312), 593–596. http://doi.org/10.1126/science.276.5312.593

Calvert, G. A., Campell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, *14*(2), 427–438. http://doi.org/10.1006/nimg.2001.0812

Calvert, G., Spence, C., & Stein, B. E. (2012). *The new handbook of multisensory processes*. *The handbook of multisensory processes*. The MIT Press. http://doi.org/nicht verfügbar?

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *Journal of Neuroscience*, *30*(38), 12572–12580. http://doi.org/10.1523/JNEUROSCI.1099-10.2010

Carp, J. (2012a). On the plurality of (methodological) worlds: estimating the analytic flexibility of fmri experiments. *Frontiers in Neuroscience*, (6), 149. http://doi.org/10.3389/fnins.2012.00149

Carp, J. (2012b). The secret lives of experiments: methods reporting in the fMRI literature. *NeuroImage*, *63*(1), 289–300. http://doi.org/10.1016/j.neuroimage.2012.07.004

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187. http://doi.org/10.1002/hbm.460020306

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT press. http://doi.org/10.1017/CBO9781107415324.004

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*(4), 495–506. http://doi.org/10.1016/S1388-2457(02)00024-X

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. http://doi.org/10.1126/science.aac4716

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique*, *30*(1), 5–19. http://doi.org/10.1016/S0987-7053(00)00055-1

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *The Journal of Neuroscience*, *36*(38), 9888–9895. http://doi.org/10.1523/JNEUROSCI.1396-16.2016

de-Wit, L., Machilsen, B., & Putzeys, T. (2010). Predictive Coding and the Neural Response to Predictable Stimuli. *Journal of Neuroscience*, *30*(26), 8702–8703. http://doi.org/10.1523/JNEUROSCI.2248-10.2010

Delorme, A., Makeig, S., & Sejnowski, T. (2001). Automatic artifact rejection for EEG data using high-order statistics and independent component analysis.

Proceedings of the Third International ICA Conference, 9–12. http://doi.org/10.1.1.11.9252

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Attention, Perception, & Psychophysics*, *66*(8), 1388–1404. http://doi.org/10.3758/BF03195006

Doesburg, S. M., Emberson, L. L., Rahi, A., Cameron, D., & Ward, L. M. (2008). Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Experimental Brain Research*, *185*(1), 11–20. http://doi.org/10.1007/s00221-007-1127-5

Doig, H. R., & Boylan, C. (1989). Presaccadic spike potentials with large horizontal eye movements. *Electroencephalography and Clinical Neurophysiology*, *73*(3), 260–263. http://doi.org/10.1016/0013-4694(89)90127-2

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *NeuroImage*, *14*(6), 1256–1267. http://doi.org/10.1006/nimg.2001.0946

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Neurophysiology*, *87*(1), 615–620. http://doi.org/10.1152/jn.00636.2001

Elbert, T., Lutzenberger, W., Rockstroh, B., & Birbaumer, N. (1985). Removal of ocular artifacts from the EEG - a biophysical approach to the EOG. *Electroencephalography and Clinical Neurophysiology*, *60*(5), 455–463. http://doi.org/10.1016/0013-4694(85)91020-X

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*(10), 704. http://doi.org/10.1038/35094565

Engel, A. K., Senkowski, D., & Schneider, T. R. (2012). *Multisensory Integration through Neural Coherence*. *The Neural Bases of Multisensory Processing*. CRC Press/Taylor & Francis. http://doi.org/NBK92855 [bookaccession]

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. *Human Body Perception from the inside out*, *131*, 105–131.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. http://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept.

Trends in Cognitive Sciences, *8*(4), 162–169. http://doi.org/10.1016/j.tics.2004.02.002

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 1–12. http://doi.org/10.1167/10.1.6

Fan, J., Byrne, J., Worden, M. S., Guise, K. G., McCandliss, B. D., Fossella, J., & Posner, M. I. (2007). The Relation of Brain Oscillations to Attentional Networks. *The Journal of Neuroscience*, *27*(23), 6197–6206. http://doi.org/10.1523/jneurosci.1833-07.2007

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*(4), 719–725. http://doi.org/10.3758/BF03194432

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013a). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442. http://doi.org/10.1017/CBO9781107415324.004

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013b). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews. Neuroscience*, *14*(6), 429–442. http://doi.org/10.1038/nrn3503

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–154. http://doi.org/10.1038/nn.2983

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, *29*(49), 15601–15612. http://doi.org/10.1523/JNEUROSCI.2574-09.2009

Fort, A., Delpuech, C., Pernier, J., & Giard, M.-H. (2002). Early auditory–visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, *14*(1), 20–30. http://doi.org/10.1016/S0926-6410(02)00058-7

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, *10*(1), 77–83. http://doi.org/10.1016/S0926-6410(00)00024-0

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, *16*(5), 419–423. http://doi.org/10.1097/00001756-200504040-00001

Foxe, J. J., Simpson, G. V, & Ahlfors, S. P. (1998). Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport*, *9*(17), 3929–3933. http://doi.org/10.1097/00001756-199812010-00030

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*. http://doi.org/10.3389/fpsyg.2011.00154

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480. http://doi.org/10.1016/j.tics.2005.08.011

Fruend, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*(6), 16. http://doi.org/10.1167/11.6.16

Fu, K. M. G., Foxe, J. J., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. *Cognitive Brain Research*, *12*(1), 145–152. http://doi.org/10.1016/S0926-6410(01)00034-9

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*(6784), 347. http://doi.org/Doi 10.1038/35012613

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203. http://doi.org/10.3758/BF03193720

Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. *American Science*, *102*(6), 460. http://doi.org/10.1511/2014.111.460

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, *25*(20), 5004–5012. http://doi.org/10.1523/JNEUROSCI.0799-05.2005

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. http://doi.org/10.1016/j.tics.2006.04.008

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*, *106*, 134–143. http://doi.org/10.1016/j.neuroimage.2014.11.036

Giard, M. H., & Besle, J. (2010). Methodological considerations: Electrophysiology of multisensory interactions in humans. In *Multisensory Object Perception in the Primate Brain* (pp. 55–70). Springer New York. http://doi.org/10.1007/978-1-4419-5615-6_4

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490. http://doi.org/10.1162/089892999563544

Gilbert, A. N., Martin, R., & Kemp, S. E. (1996). Cross-modal correspondence between vision and olfaction: the color of smells. *American Journal of Psychology*, 335–351.

Glenberg, A. M., & Fernandez, A. (1988). Evidence for auditory temporal distinctiveness: modality effects in order and frequency judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(4), 728. http://doi.org/10.1037/0278-7393.14.4.728

Glenberg, A. M., Mann, S., Altman, L., Forman, T., & Procise, S. (1989). Modality effects in the coding and reproduction of rhythms. *Memory & Cognition*, *17*(4), 373–383. http://doi.org/10.3758/BF03202611

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998a). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *47*(6), 1464–1480.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998b). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *47*(6), 1464–1480.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. http://doi.org/10.1371/journal.pbio.1001752

Gu, Y., Angelaki, D. E., & Deangelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, *11*(10), 1201–1210. http://doi.org/10.1038/nn.2191

Händel, B. F., Haarmeier, T., & Jensen, O. (2011). Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of Cognitive Neuroscience*, *23*(9), 2494–2502. http://doi.org/10.1162/jocn.2010.21557

Hecht, D., Reiner, M., & Karni, A. (2008a). Enhancement of response times to bi- and tri-modal sensory stimuli during active movements. *Experimental Brain Research*, *185*(4), 655. http://doi.org/10.1007/s00221-007-1191-x

Hecht, D., Reiner, M., & Karni, A. (2008b). Multisensory enhancement: gains in

choice and in simple response times. *Experimental Brain Research*, *189*(2), 133–143. http://doi.org/10.1007/s00221-008-1410-0

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, *179*(4), 595–606. http://doi.org/10.1007/s00221-006-0814-y

Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., … Noppeney, U. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. *NeuroImage*, *60*(2), 1063–1072. http://doi.org/10.1016/j.neuroimage.2011.09.072

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1–1. http://doi.org/10.1167/4.12.1

Hipp, J., & Siegel, M. (2013). Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Frontiers in Human Neuroscience*. Frontiers. http://doi.org/10.3389/fnhum.2013.00338

Hoffmann, S., & Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PLoS ONE*, *3*(8), e3004. http://doi.org/10.1371/journal.pone.0003004

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, *49*(1), 875–884. http://doi.org/10.1016/j.neuroimage.2009.08.058

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow & Metabolism*, *16*(1), 7–22. http://doi.org/10.1097/00004647-199601000-00002

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. http://doi.org/10.1002/wcs.142

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2016). A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573. http://doi.org/10.1002/hbm.23471

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., Van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The deceptively simple N170 reflects network information processing mechanisms involving visual feature coding and transfer across hemispheres. *Cerebral Cortex*, *26*(11), 4123–4135. http://doi.org/10.1093/cercor/bhw196

Ince, R. A. A., Panzeri, S., & Kayser, C. (2013). Neural codes formed by small

and temporally precise populations in auditory cortex. *Journal of Neuroscience*, *33*(46), 18277–18287. http://doi.org/10.1523/JNEUROSCI.2631-13.2013

Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., & Artieda, J. (2003). Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. *Journal of Clinical Neurophysiology*, *20*(4), 249–257. http://doi.org/10.1097/00004691-200307000-00004

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200–206. http://doi.org/10.1016/j.tics.2012.03.002

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscienceeuroscience*, *15*(8), 1395–1399. http://doi.org/10.1046/j.1460-9568.2002.01975.x

Kaiser, J., Hertrich, I., Ackermann, H., & Lutzenberger, W. (2006). Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. *NeuroImage*, *30*(4), 1376–1382. http://doi.org/10.1016/j.neuroimage.2005.10.042

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2004). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex*, *15*(5), 646–653. http://doi.org/10.1093/cercor/bhh166

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current Biology*, *16*(11), 1096–1102. http://doi.org/10.1016/j.cub.2006.04.003

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function*, *212*(2), 121–132. http://doi.org/10.1007/s00429-007-0154-0

Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, *61*(4), 597–608. http://doi.org/10.1016/j.neuron.2009.01.008

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, *48*(2), 373–384. http://doi.org/10.1016/j.neuron.2005.09.018

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, *27*(8), 1824–1835. http://doi.org/10.1523/JNEUROSCI.4737-06.2007

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, *18*(7), 1560–1574. http://doi.org/10.1093/cercor/bhm187

Kayser, C., & Shams, L. (2015). Multisensory Causal Inference in the Brain. *PLOS Biology*, *13*(2), e1002075. http://doi.org/10.1371/journal.pbio.1002075

Kayser, S. J., Ince, R. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *The Journal of Neuroscience*, *35*(44), 14691–14701. http://doi.org/10.1523/JNEUROSCI.2243-15.2015

Kayser, S. J., McNair, S. W., & Kayser, C. (2016). Prestimulus influences on auditory perception from sensory representations and decision processes. *Proceedings of the National Academy of Sciences*, *113*(17), 4842–4847. http://doi.org/10.1073/pnas.1524087113

Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage*, *147*, 32–42. http://doi.org/10.1016/j.neuroimage.2016.11.062

Keller, A., Payne, L., & Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia*, *99*, 48–63. http://doi.org/10.1016/j.neuropsychologia.2017.02.021

Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. *NeuroImage*, *49*(3), 2248–2263. http://doi.org/10.1016/j.neuroimage.2009.10.057

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. http://doi.org/10.1016/j.tics.2012.10.007

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, *244*(2), 73–76. http://doi.org/10.1016/S0304-3940(98)00122-0

Knill, D. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7), 5–5. http://doi.org/10.1167/7.7.5

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. http://doi.org/10.1016/j.tins.2004.10.007

Knill, D. C., & Saunders, J. a. (2003). Do humans optimally integrate stereo and

texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. http://doi.org/10.1016/S0042-6989(03)00458-9

Kohler, W. (1929). *Gestalt Psychology*. *New York : H. Liveright*. New York: H. Liveright.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9), e943. http://doi.org/10.1371/journal.pone.0000943

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. http://doi.org/10.1016/j.cognition.2009.08.016

Krug, K., Cicmil, N., Parker, A. J., & Cumming, B. G. (2013). A causal role for V5/MT neurons coding motion-disparity conjunctions in resolving perceptual ambiguity. *Current Biology*, *23*(15), 1454–1459. http://doi.org/10.1016/j.cub.2013.06.023

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–292. http://doi.org/10.1016/j.neuron.2006.12.011

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y.-F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14*(3), 420–429. http://doi.org/10.1162/089892902317361930

Lehmann, C., Herdener, M., Esposito, F., Hubl, D., di Salle, F., Scheffler, K., … Seifritz, E. (2006). Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *NeuroImage*, *31*(1), 294–300. http://doi.org/10.1016/j.neuroimage.2005.12.038

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: An experimental investigation. *Knowledge and Information Systems*, *10*(4), 453–472. http://doi.org/10.1007/s10115-006-0013-y

Li, Y., Ma, Z., Lu, W., & Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiological Measurement*, *27*(4), 425–436. http://doi.org/10.1088/0967-3334/27/4/008

Lou, B., Li, Y., Philiastides, M. G., & Sajda, P. (2014). Prestimulus alpha power predicts fidelity of sensory encoding in perceptual decision making. *NeuroImage*, *87*, 242–251. http://doi.org/10.1016/j.neuroimage.2013.10.041

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light

enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453. http://doi.org/10.1016/S0926-6410(03)00160-5

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes mechanisms of attention. *Vision Research*, *38*(9), 1183–1198. http://doi.org/10.1016/B978-012375731-9/50078-1

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. http://doi.org/10.1111/psyp.12639

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), 25–26. http://doi.org/10.1371/journal.pbio.1000445

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of Bimodal Looming Signals through Neuronal Coherence in the Temporal Lobe. *Current Biology*, *18*(13), 963–968. http://doi.org/10.1016/j.cub.2008.05.043

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. http://doi.org/10.1016/j.neuron.2008.06.024

Markett, S., Reuter, M., Montag, C., Voigt, G., Lachmann, B., Rudorf, S., … Weber, B. (2014). Assessing the function of the fronto-parietal attention network: Insights from resting-state fMRI and the attentional network test. *Human Brain Mapping*, *35*(4), 1700–1709. http://doi.org/10.1002/hbm.22285

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. In *International Journal of Psychophysiology* (Vol. 50, pp. 125–145). http://doi.org/10.1016/S0167-8760(03)00129-6

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *Perception*, *29*(6), 745–754. http://doi.org/10.1068/p2984

Matsuhashi, M., Ikeda, A., Ohara, S., Matsumoto, R., Yamamoto, J., Takayama, M., … Mikuni, N. (2004). Multisensory convergence at human temporo-parietal junction-epicortical recording of evoked responses. *Clinical Neurophysiology*, *115*(5), 1145–1160. http://doi.org/10.1016/j.clinph.2003.12.009

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. http://doi.org/10.1038/264746a0

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically

corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323. http://doi.org/10.1037/0096-3445.116.4.323

Meltzer, J. A., Zaveri, H. P., Goncharova, I. I., Distasio, M. M., Papademetris, X., Spencer, S. S., … Constable, R. T. (2008). Effects of working memory load on oscillatory power in human intracranial EEG. *Cerebral Cortex*, *18*(8), 1843–1855. http://doi.org/10.1093/cercor/bhm213

Meredith, M. A., & Stein, B. E. (1983a). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. http://doi.org/10.1126/science.6867718

Meredith, M. A., & Stein, B. E. (1983b). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. http://doi.org/10.1126/science.6867718

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of muitisensory neurons in cat superior colliculus. *Brain Research*, *5*, 350–354. http://doi.org/10.1016/0006-8993(86)91648-3

Meredith, M. A., & Stein, B. E. (1986a). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. http://doi.org/citeulike-article-id:844215

Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. http://doi.org/citeulike-article-id:844215

Molholm, S. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, *96*(2), 721–729. http://doi.org/10.1152/jn.00285.2006

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, *14*(1), 115–128. http://doi.org/10.1016/S0926-6410(02)00066-6

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133–136. http://doi.org/10.3758/CABN.4.2.133

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163. http://doi.org/10.1016/S0926-6410(03)00089-2

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory

integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–673. http://doi.org/10.1016/j.neuron.2008.06.024

Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*. http://doi.org/10.1038/srep18253

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, *13*(3), 417–425. http://doi.org/10.1016/S0926-6410(02)00053-8

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters*, *363*(2), 112–115. http://doi.org/10.1016/j.neulet.2004.03.076

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., … Foxe, J. J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, *15*(7), 963–974. http://doi.org/10.1093/cercor/bhh197

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience*, *7*. http://doi.org/10.3389/fnhum.2013.00138

Nichols, T. E., & Holmes, A. P. (2001). Nonparametric Permutation Tests for functional Neuroimaging Experiments: A Primer with examples. *Human Brain Mapping*, *15*(1), 1–25. http://doi.org/10.1002/hbm.1058

Nuijten, M. B., Hartgerink, C. H. J., Assen, M. A. L. M. van, Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. http://doi.org/10.3758/s13428-015-0664-2

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. http://doi.org/10.1155/2011/156869

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468. http://doi.org/10.1016/S0042-6989(03)00435-8

Ostwald, D., & Bagshaw, A. P. (2011). Information theoretic approaches to functional neuroimaging. *Magnetic Resonance Imaging*, *29*(10), 1417–1428. http://doi.org/10.1016/j.mri.2011.07.013

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal

ventriloquism effect. *Neuroscience Letters*, *442*(3), 257–261. http://doi.org/10.1016/j.neulet.2008.07.010

Parise, C. V., & Spence, C. (2009). "When birds of a feather flock together": synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, *4*(5), e5664. http://doi.org/10.1371/journal.pone.0005664

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333. http://doi.org/10.1007/s00221-012-3140-6

Parise, C. V, Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108. http://doi.org/10.1073/pnas.1322705111

Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., & Sajda, P. (2002). Linear Spatial Integration for Single-Trial Detection in Encephalography. *NeuroImage*, *17*, 223–230. http://doi.org/10.1006/nimg.2002.1212

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*(2), 326–341. http://doi.org/10.1016/j.neuroimage.2005.05.032

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience*, *31*(10), 1783–1790. http://doi.org/10.1111/j.1460-9568.2010.07193.x

Paulus, M. P., Hozack, N., Zauscher, B., McDowell, J. E., Frank, L., Brown, G. G., & Braff, D. L. (2001). Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty. *NeuroImage*, *13*(1), 91–100. http://doi.org/10.1006/nimg.2000.0667

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., & Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Human Brain Mapping*, *27*(6), 471–477. http://doi.org/10.1002/hbm.20190

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*, *16*(2), 125–128. http://doi.org/10.1097/00001756-200502080-00010

Pernet, C. R., Sajda, R., & Rousselet, G. A. (2011). Single-trial analyses: why bother? *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2011.00322

Philiastides, M. G., Heekeren, H. R., & Sajda, P. (2014). Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. *Journal of Neuroscience*, *34*(50), 16877–16889. http://doi.org/10.1523/JNEUROSCI.3012-14.2014

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*(4), 509–518. http://doi.org/10.1093/cercor/bhi130

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(May), 15808. http://doi.org/10.1038/ncomms15808

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*. http://doi.org/10.3389/fnhum.2012.00278

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., … Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *The Journal of Neuroscience*, *21*(9), 3175–3183. http://doi.org/21/9/3175 [pii]

Ramachandran, V. S., & Hubbard, E. . (2001). Synaesthesia: a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience*, *17*(12), 1784–1792. http://doi.org/10.1038/nn.3865

Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience*, *32*(11), 3726–3735. http://doi.org/10.1523/JNEUROSCI.4998-11.2012

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6539–6544. http://doi.org/10.1073/pnas.0812589106

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, *89*(2), 1078–1093. http://doi.org/10.1152/jn.00706.2002

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1085–1099. http://doi.org/10.1037/0096-1523.28.5.1085

Repp, B. H., & Penel, A. (2002). Auditory Dominance in Temporal Processing: New Evidence From Synchronization with Simultaneous Visual and Auditory Sequences. *Journal of Experimental Psychology: Human Perception and Performance.*, *28*(5), 1085–1099.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1598), 2159–2168.

Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, *43*(6), 1255–1259. http://doi.org/10.1121/1.1910976

Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2015). Statistically optimal multisensory cue integration: a practical tutorial. *Multisensory Research*, *29*(4–5), 279–317. http://doi.org/10.1163/22134808-00002510

Rohe, T., & Noppeney, U. (2015a). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073. http://doi.org/10.1371/journal.pbio.1002073

Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes Bayesian Causal Inference in perception via two mechanisms. *Journal of Vision*, *15*(5), 22–22. http://doi.org/10.1167/15.5.22.doi

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, *26*(4), 509–514. http://doi.org/10.1016/j.cub.2015.12.056

Rohenkohl, G., & Nobre, A. C. (2011). Alpha Oscillations Related to Anticipatory Attention Follow Temporal Expectations. *Journal of Neuroscience*, *31*(40), 14076–14084. http://doi.org/10.1523/JNEUROSCI.3387-11.2011

Rombouts, S. A. R. B., Keunen, R. W. M., & Stam, C. J. (1995). Investigation of nonlinear structure in multidimensional EEG. *Physical Letters A*, *202*(July), 352–358. http://doi.org/10.1016/0375-9601(95)00335-Z

Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *The Journal of Neuroscience*, *30*(25), 8692–8697. http://doi.org/10.1523/JNEUROSCI.0160-10.2010

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A*, *22*(5), 801–809. http://doi.org/10.1364/JOSAA.22.000801

Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion

in slant discrimination: failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, *7*(6), 3–3. http://doi.org/10.1167/7.6.3

Rosenthal, R. (1994). Parametric measures of effect size. In L. V. Hedges & H. M. Cooper (Eds.), *The Handbook of Research Synthesis* (pp. 231–243). New York: Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (2003). requivalent: A simple Effect Size Indicator. *Psychological Methods*, *8*(4), 492–496. http://doi.org/10.1037/1082-989X.8.4.492

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*(1), 16–25. http://doi.org/10.1016/j.tics.2013.10.010

Rupp, A., Hack, S., Gutschalk, A., Schneider, P., Picton, T. W., Stippich, C., & Scherg, M. (2000). Fast temporal interactions in human auditory cortex. *Neuroreport*, *11*(17), 3731–3736.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113–129. http://doi.org/10.1016/j.cognition.2005.01.004

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *The Journal of Neuroscience*, *29*(20), 6490–6499. http://doi.org/10.1523/JNEUROSCI.5437-08.2009

Sajda, P., Philiastides, M. G., & Parra, L. C. (2009). Single-Trial Analysis of Neuroimaging Data: Inferring Neural Networks Underlying Perceptual Decision-Making in the Human Brain. *IEEE Reviews in Biomedical Engineering*, *2*, 97–109. http://doi.org/10.1109/RBME.2009.2034535

Sakowitz, O. W., Quiroga, R. Q., Schurmann, M., & Ba Ar, E. (2001). Bisensory stimulation increases gamma-responses over multiple cortical regions. *Cognitive Brain Research*, *11*(2), 267–279. http://doi.org/10.1016/S0926-6410(00)00081-1

Sakowitz, O. W., Quiroga, R. Q., Schurmann, M., & Basar, E. (2005). Spatio-temporal frequency characteristics of intersensory components in audiovisually evoked potentials. *Cognitive Brain Research*, *23*(2), 316–326. http://doi.org/10.1016/j.cogbrainres.2004.10.012

Sakowitz, O. W., Schürmann, M., & Başar, E. (2000). Oscillatory frontal theta responses are increased upon bisensory stimulation. *Clinical Neurophysiology*, *111*(5), 884–893. http://doi.org/10.1016/S1388-2457(99)00315-6

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., … Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*(11), 2917–2926. http://doi.org/10.1111/j.1460-9568.2005.04482.x

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., & Bastiaansen, M. C. M. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *NeuroImage*, *44*(3), 1224–1238. http://doi.org/10.1016/j.neuroimage.2008.08.041

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage*, *42*(3), 1244–1254. http://doi.org/10.1016/j.neuroimage.2008.05.033

Schroeder, C. E., Mehta, A. D., & Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cerebral Cortex*, *8*(7), 575–92. http://doi.org/10.1093/cercor/8.7.575

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*(14), 1422–1427. http://doi.org/10.1016/j.cub.2006.05.048

Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2006). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. *Cerebral Cortex*, *16*(11), 1556–1565. http://doi.org/10.1093/cercor/bhj091

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. http://doi.org/10.1016/j.neuroimage.2011.03.075

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*. http://doi.org/10.1016/j.tins.2008.05.002

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. http://doi.org/10.1016/j.tics.2010.07.001

Shams, L., Ji Ma, W., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–1927.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788. http://doi.org/10.1038/35048669

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*(1), 147–152. http://doi.org/10.1016/S0926-6410(02)00069-1

Sheppard, J. P., Raposo, D., & Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *Journal of Vision*, *13*(6), 4–4. http://doi.org/10.1167/13.6.4

Shomstein, S., Kravitz, D. J., & Behrmann, M. (2012). Attentional control: Temporal relationships within the fronto-parietal network. *Neuropsychologia*, *50*(6), 1202–1210. http://doi.org/10.1016/j.neuropsychologia.2012.02.009

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, *13*(2), 121. http://doi.org/10.1038/nrn3137

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, *39*(4), 553–569. http://doi.org/10.1068/p6591

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention Perception & Psychophysics*, *73*(4), 971–995. http://doi.org/10.3758/s13414-010-0073-7

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*. http://doi.org/10.1016/j.concog.2012.12.006

Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Frontiers in Integrative Neuroscience*, *3*. http://doi.org/10.3389/neuro.07.002.2009

Stein, B. E., & Meredith, M. E. (1993). *The Merging of the Senses*. The MIT Press. http://doi.org/10.3389/neuro.01.019.2008

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, *198*(2–3), 113–126. http://doi.org/10.1007/s00221-009-1880-8

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, *54*(2), 407–411. http://doi.org/10.1073/pnas.54.2.407

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223. http://doi.org/10.1016/j.neuroimage.2008.09.034

Stevenson, R. A., Kim, S., & James, T. W. (2009). An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Experimental Brain Research*, *198*(2–3), 183–194. http://doi.org/10.1007/s00221-009-1783-8

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. http://doi.org/10.1016/j.tics.2010.06.008

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114. http://doi.org/10.1162/0898929054475172

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106–114. http://doi.org/10.1016/S0926-6410(02)00065-4

Thickbroom, G. W., & Mastaglia, F. L. (1986). Presaccadic spike potential. Relation to eye movement direction. *Electroencephalography and Clinical Neurophysiology*, *64*(3), 211–214. http://doi.org/10.1016/0013-4694(86)90167-7

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *Journal of Neuroscience*, *31*(10), 3853–3861. http://doi.org/10.1523/JNEUROSCI.6176-10.2011

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, *26*(37), 9494–9502. http://doi.org/10.1523/JNEUROSCI.0875-06.2006

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, *24*(11), 1256–1262. http://doi.org/10.1016/j.cub.2014.04.020

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, *120*(5), 868–877. http://doi.org/10.1016/j.clinph.2009.01.015

Vroomen, J. ;, & De Gelder, B. (2000). Sound Enhances Visual Perception: Cross-Modal Effects of Auditory Organization on Vision. *Journal of Experimental Psychology:Human Perception and Performance*, *26*(5), 1583. http://doi.org/10.1037//0096-1523.26.5.1583

Vroomen, J., Bertelson, P., & de Gelder, B. (1998). A visual influence in the discrimination of auditory location. *AVSP'98 Proceedings of the International Conference on Auditory-Visual Speech Processing*.

Vroomen, J., & De Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, *90*(3), 1195–268. http://doi.org/10.1152/physrev.00035.2008

Werner, S., & Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, *20*(8), 1829–1842. http://doi.org/10.1093/cercor/bhp248

Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., & Schröger, E. (2007). Binding symbols and sounds: Evidence from event-related oscillatory gamma-band activity. *Cerebral Cortex*, *17*(11), 2696–2702. http://doi.org/10.1093/cercor/bhl178

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., & Schröger, E. (2004). From symbols to sounds: visual symbolic information activates sound representations. *Psychophysiology*, *41*(5), 709–715. http://doi.org/10.1111/j.1469-8986.2004.00208.x

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci*, *10*(6), 685–686. http://doi.org/10.1038/nn1900

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *The Journal of Neuroscience*, *20*(6), 1–6. http://doi.org/Rc63

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron*, *76*(4), 847–858. http://doi.org/10.1016/j.neuron.2012.09.015

Yamagishi, N., Callan, D. E., Goda, N., Anderson, S. J., Yoshida, Y., & Kawato, M. (2003). Attentional modulation of oscillatory activity in human visual cortex. *NeuroImage*, *20*(1), 98–113. http://doi.org/10.1016/S1053-8119(03)00341-0

Yamagishi, N., Goda, N., Callan, D. E., Anderson, S. J., & Kawato, M. (2005). Attentional shifts towards an expected visual target alter the level of alpha-band oscillatory activity in the human calcarine cortex. *Cognitive Brain Research*, *25*(3), 799–809. http://doi.org/10.1016/j.cogbrainres.2005.09.006

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal frequency channels are linked across audition and touch. *Current Biology*, *19*(7), 561–566. http://doi.org/10.1016/j.cub.2009.02.013

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience*, *27*(5), 1090–1096. http://doi.org/10.1523/JNEUROSCI.4828-06.2007